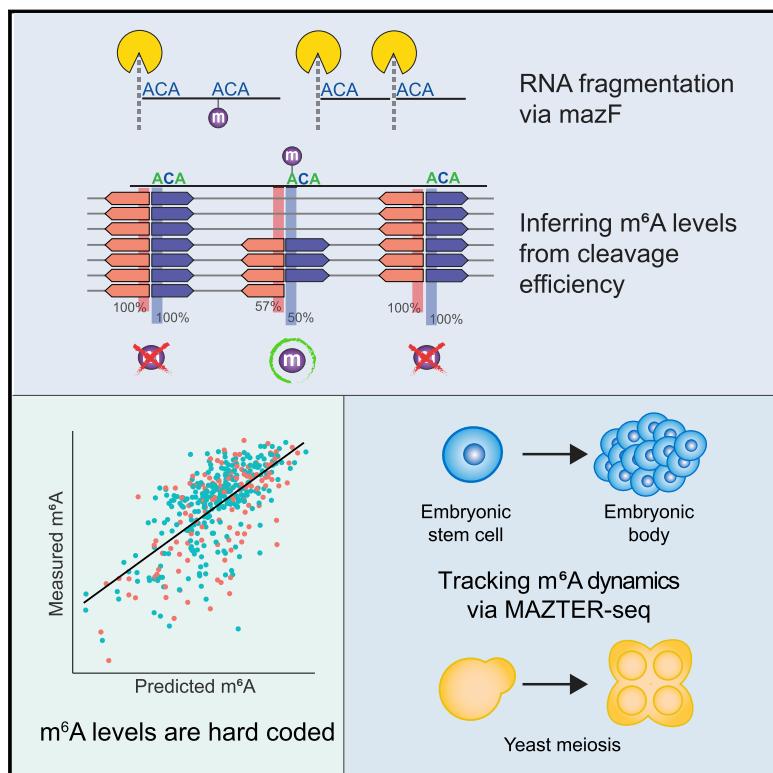


Deciphering the “m⁶A Code” via Antibody-Independent Quantitative Profiling

Graphical Abstract



Authors

Miguel Angel Garcia-Campos,
Sarit Edelheit, Ursula Toth, ...,
Jacob H. Hanna, Walter Rossmanith,
Schraga Schwartz

Correspondence

schwartz@weizmann.ac.il

In Brief

A new enzymatic approach for precise mapping and measurement of m⁶A within mRNAs provides insight into how methylation sites are selected and the functional impact of the modifications.

Highlights

- RNA digestion via m⁶A sensitive RNase (MAZTER-seq) allows systematic m⁶A quantitation
- MAZTER-seq reveals that antibody-based methods are of limited sensitivity
- m⁶A stoichiometry is “hard coded” by a simple, predictable, and conserved code
- MAZTER-seq allows quantitative tracking of m⁶A in diverse biological settings



Deciphering the “m⁶A Code” via Antibody-Independent Quantitative Profiling

Miguel Angel Garcia-Campos,^{1,4} Sarit Edelheit,^{1,4} Ursula Toth,² Modi Safra,¹ Ran Shachar,¹ Sergey Viukov,¹ Roni Winkler,¹ Ronit Nir,¹ Lior Lasman,¹ Alexander Brandis,³ Jacob H. Hanna,¹ Walter Rossmanith,² and Schraga Schwartz^{1,5,*}

¹Department of Molecular Genetics, Weizmann Institute of Science, 7610001 Rehovot, Israel

²Medical University of Vienna, Center for Anatomy & Cell Biology, Währinger Straße 13, 1090 Vienna, Austria

³Life Sciences Core Facilities, Weizmann Institute of Science, 7610001 Rehovot, Israel

⁴These authors contributed equally

⁵Lead Contact

*Correspondence: schwartz@weizmann.ac.il

<https://doi.org/10.1016/j.cell.2019.06.013>

SUMMARY

N6-methyladenosine (m⁶A) is the most abundant modification on mRNA and is implicated in critical roles in development, physiology, and disease. A major limitation has been the inability to quantify m⁶A stoichiometry and the lack of antibody-independent methodologies for interrogating m⁶A. Here, we develop MAZTER-seq for systematic quantitative profiling of m6A at single-nucleotide resolution at 16%–25% of expressed sites, building on differential cleavage by an RNase. MAZTER-seq permits validation and *de novo* discovery of m⁶A sites, calibration of the performance of antibody-based approaches, and quantitative tracking of m⁶A dynamics in yeast gametogenesis and mammalian differentiation. We discover that m6A stoichiometry is “hard coded” in *cis* via a simple and predictable code, accounting for 33%–46% of the variability in methylation levels and allowing accurate prediction of m⁶A loss and acquisition events across evolution. MAZTER-seq allows quantitative investigation of m⁶A regulation in subcellular fractions, diverse cell types, and disease states.

INTRODUCTION

m⁶A is the most abundant modification on mRNA. Although discovered nearly five decades ago, the inability to map m⁶A on mRNA imposed strong limitations for functionally dissecting its roles. Seven years ago, we and others developed immunoprecipitation-based approaches coupled with high-throughput sequencing (m⁶A-seq, m⁶A-MeRIP), allowing to detect regions harboring m⁶A (“m⁶A peaks”) (Dominissini et al., 2012; Meyer et al., 2012). These approaches paved the way to major advances in the understanding of m⁶A, its distribution, and conservation and have facilitated the functional and mechanistic dissection of m⁶A in development and disease (reviewed in

Knuckles and Bübler, 2018; Meyer and Jaffrey, 2017; Schwartz, 2016; Yue et al., 2015). Nonetheless, antibody-based approaches have several important limitations. First, antibody-based approaches are of limited utility for quantification of m⁶A stoichiometry. Resolving m⁶A stoichiometry is critical for functional prioritization of m⁶A sites and for addressing critical questions pertaining to the biogenesis, regulation, and dynamics of m⁶A within cells and across stimuli (Grozhik and Jaffrey, 2018; Meyer and Jaffrey, 2014; Schwartz, 2016). Second, the specificity of detection of m⁶A sites is limited due to promiscuous binding (Schwartz et al., 2013) and to cross-reactivity with related modifications, such as m⁶Am (Dominissini et al., 2012; Linder et al., 2015; Schwartz et al., 2014a). Third, the sensitivity of m⁶A detection using antibodies could thus far not be evaluated in the absence of an orthogonal technique allowing independent systematic profiling of m⁶A. Additional limitations of classical m⁶A-seq approaches are their limited resolution and the requirement for large amounts of starting material (Dominissini et al., 2012; Meyer et al., 2012). While the resolution can be enhanced through crosslinking of the anti-m⁶A antibody to the RNA (Ke et al., 2015; Linder et al., 2015) and required starting amounts have been decreased in recent protocol optimizations (Merkurjev et al., 2018; Zeng et al., 2018), no protocol exists providing complex, single-nucleotide resolution mapping from limited starting mRNA material.

The inability to quantify m⁶A stoichiometry at individual sites has severely limited our ability to address critical questions pertaining to the distribution, dynamics, and function of m⁶A. A key, poorly understood question pertains to the code governing the presence of m⁶A at some sites, but not at others (the “m⁶A code”) (Darnell et al., 2018; Meyer and Jaffrey, 2017). Specifically, the sequence motif at which m⁶A is present in mammals often represented as DRACH (D = A/G/U, R = A/G, H = A/C/U) is ~20-fold more abundant than the number of m⁶A sites typically identified in an m⁶A mapping experiment. Moreover, the stoichiometry of m6A at distinct sites varies substantially (Horowitz et al., 1984; Liu et al., 2013). What governs deposition of m⁶A? Is there a single, global code governing m⁶A deposition? Or is each site subjected to a distinct, complex modulatory regime through factors regulating deposition, decoding, and



removal (Fu et al., 2014; Wang et al., 2014; Zhao et al., 2018)? Distinguishing between these two scenarios is of the utmost importance for understanding the key determinants underlying m⁶A levels, the potential for dynamic modulation of m⁶A, and the constraints on the evolution of new m⁶A sites. In addition, lack of quantitative measurements has further limited our ability to quantitatively interrogate the impact of m⁶A on the molecular fate of mRNA. While this has been an area of intense research (Knuckles and Bübler, 2018; Meyer and Jaffrey, 2017; Yue et al., 2015), to date, investigations of these aspects have mostly been qualitative in nature, rendering it difficult to infer direct, dose-dependent causal effects. Finally, the absence of quantitative techniques has severely limited our ability to dissect the extent to which m⁶A is dynamically modulated within different subcellular compartments and across stimuli, responses, and disease states.

Here, we build on the ability of the MazF RNase to cleave RNA at unmethylated sites occurring at ACA motifs, but not at their methylated counterparts (Imanishi et al., 2017), and develop MAZTER-seq for quantitative, single-nucleotide detection of m⁶A at 16%–25% of all methylation sites. We establish the ability of MAZTER-seq for *de novo* detection of m⁶A, for calibrating the sensitivity and specificity of antibody-based approaches for monitoring m⁶A, for tracking of m⁶A dynamics in yeast and mammalian systems, and for interrogating m⁶A function. We reveal that deposition of m⁶A is “hard coded” primarily through the extended sequence surrounding the modified site. Changes in sequence composition, as occur naturally throughout evolution or as induced experimentally by us, lead to predictable changes in methylation levels. We anticipate that MAZTER-seq will pave the path toward quantitative investigation of m⁶A regulation in a wide variety of additional systems, including diverse cell types, stimulations, subcellular compartments, and disease states.

RESULTS

MAZTER-seq relies on the ability of the bacterial RNase MazF to cleave RNA immediately upstream of an “ACA” sequence, but not upstream of “m⁶A-CA” (Imanishi et al., 2017). MAZTER-seq combines the following steps: (1) digestion of mRNA with MazF, (2) end repair and ligation of an adaptor to the 3' of the resultant RNA fragments, (3) reverse transcription primed from the ligated adaptor, (4) ligation of a second adaptor to the 3' of the cDNA, and (5) cDNA amplification by PCR followed by paired-end sequencing (Figure 1A, top). In an idealized scenario, following MazF treatment, each fragment should begin with an ACA site (5' ACA) and terminate immediately prior to a downstream ACA site (3' ACA). Each pair of sequencing reads—which together capture the precise start and end of the original RNA fragment—provides an indication that two ACA sites (at the 5' and 3' of the interval spanning the read pair) were unmethylated in one molecule. An m⁶A-containing site is anticipated to be characterized by an abundance of reads passing through, but not terminating, at it (Figure 1A, bottom). To identify and quantify methylation sites, we developed MAZTER-MINE, a computational pipeline that quantifies the number of reads that begin, terminate, and read through each transcriptomic ACA site (Figure S1).

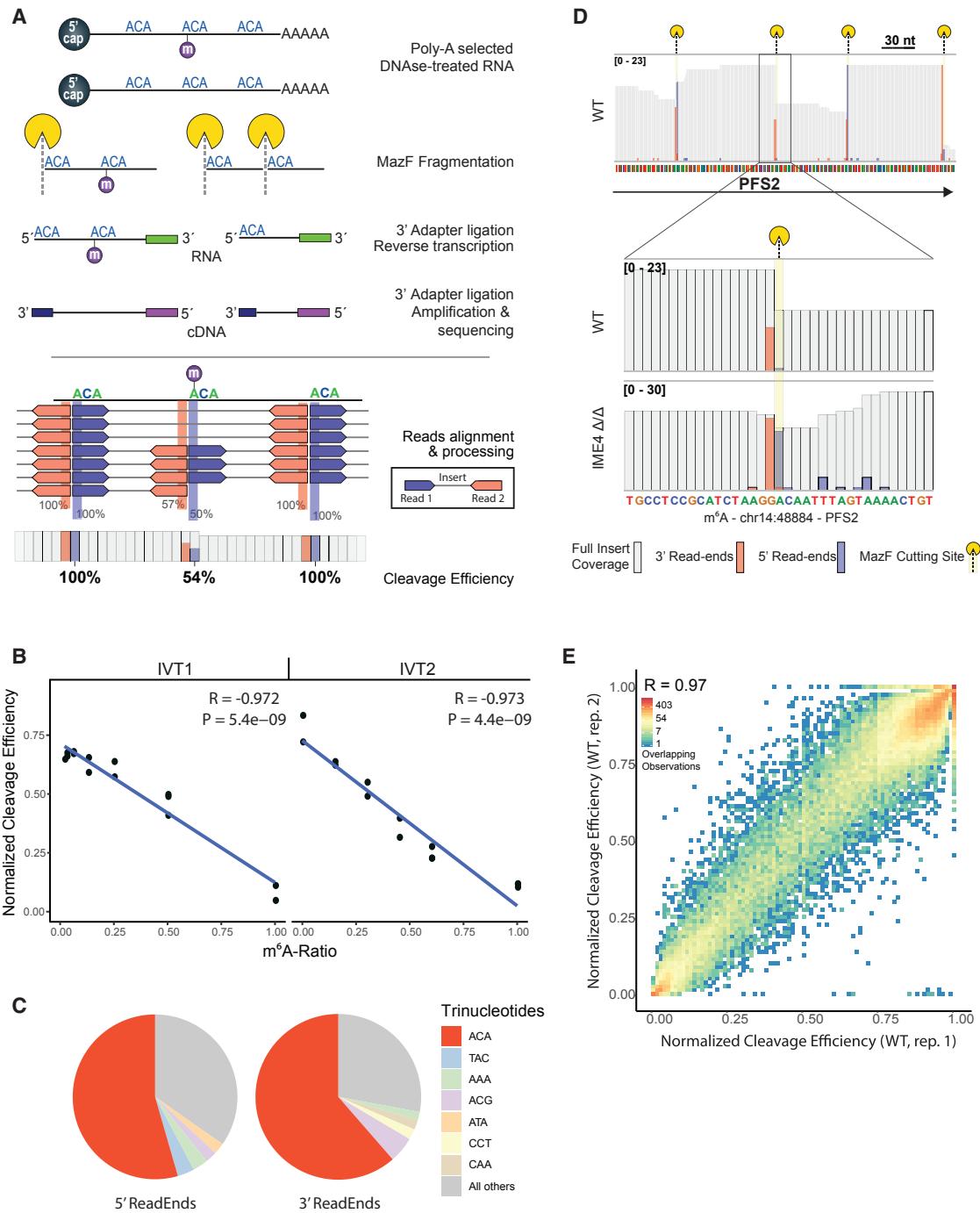
Each site is assigned a 5' and a 3' cleavage efficiency, quantifying the number of reads beginning at or ending immediately before each ACA site divided by the number of reads overlapping the site, respectively (Figure 1A, bottom). In addition, a mean cleavage efficiency metric (“cleavage efficiency”) is calculated as a mean of these two cleavage efficiencies. This pipeline pre-filters ACA sites that cannot be accurately quantified, when the distance between the interrogated ACA and its upstream and downstream ACA sites is too short, using an inferred, library-specific threshold (Figures S2A and S2B; STAR Methods). In yeast, 113,014 out of 226,058 ACA sites (50%) are considered quantifiable by MAZTER-MINE. MAZTER-MINE is available at https://github.com/SchwartzLab/mazter_mine.

To assess the potential of MAZTER-seq to quantitatively capture methylation levels, we applied MAZTER-seq to two synthetic short (88 nt long) RNA molecules harboring a single methylation site within a MazF consensus sequence (a single ACA or m⁶A-CA sequence), which were spiked into complex RNA samples at varying stoichiometries. We obtained excellent agreement between the generated m⁶A stoichiometries and the experimentally derived cleavage efficiencies ($R = -0.97$) (Figure 1B), demonstrating the quantitative power of MAZTER-seq under idealized and controlled settings. When applied to the entire yeast transcriptome (see below), 50%–60% of sequenced reads began at ACA sites, and a similar percentage terminated immediately prior to ACA sites, consistent with our expectations (Figure 1C). Mapping of the sequencing reads therefore resulted in sharp pile-ups of reads beginning at an ACA sequence and terminating immediately before the next ACA sequence (Figure 1D). Finally, we observed that the cleavage efficiency scores were highly reproducible across biological replicates ($R = 0.97$; Figure 1E).

MAZTER-seq is limited to quantifying m⁶A sites occurring at ACA motifs. All methylated adenosines are invariably followed by a C in both human and yeast (Dominissini et al., 2012; Harper et al., 1990; Horowitz et al., 1984; Meyer et al., 2012; Wei and Moss, 1977). In yeast, the most prevalent nucleotide at position +2 with respect to the modified position is an “A,” present in ~50% of all methylation sites (Schwartz et al., 2013). In mammalian systems, A is the second-most prevalent nucleotide at this position, present at roughly one-third of the detected m⁶A sites (Linder et al., 2015). The requirement for appropriately distanced ACA sites, along with the requirement for an ACA motif, thus renders ~25% of all m⁶A sites in yeast (and ~16% in mammals) amenable to quantification via MAZTER-MINE. MAZTER-seq is thus applicable to a considerable subset of methylation sites but does not allow global mapping of m⁶A.

MAZTER-Seq Allows Detecting and Quantifying m⁶A Levels at Endogenous Sites

We next tested the ability of MAZTER-seq to assay m⁶A levels at endogenous sites in yeast undergoing meiosis. Yeast mRNAs lack m⁶A under vegetative growth conditions but acquire a widespread methylation program that is catalyzed by the methyltransferase Ime4 and peaks at prophase (Agarwala et al., 2012; Clancy et al., 2002; Schwartz et al., 2013). We applied MAZTER-seq to mRNA originating from wild-type (WT) and ime4Δ/Δ strains, both in the background of the ndt80Δ/Δ

**Figure 1. Establishing MAZTER-Seq**

(A) (Top) Outline of experimental procedure. (Bottom) Computational approach. Following alignment, read starts and ends, and coverages are calculated based on which cleavage efficiencies are calculated.

(B) Correlation between m^6A stoichiometries in synthetic RNAs (x axis) and MAZTER-seq measurements (y axis).

(C) Relative frequency of the most frequent trinucleotides at the 5' and 3' termini of aligned reads.

(D) Alignment patterns at an m^6A -detected methylation site in WT and $ime4\Delta/\Delta$ cells in prophase. The number of sequence tags beginning, ending, and overlapping each site are depicted in blue, red, and white, respectively. The MazF cleaved site is highlighted in yellow.

(E) Cleavage efficiencies between replicates are highly reproducible. The color gradient depicts the density of overlapping points.

deletion that synchronizes yeast cells at meiotic prophase (Brar et al., 2012). MAZTER-seq was applied to the mRNA (“Input”), following which—as an additional control—we subjected part of the Input mRNA to m⁶A-immunoprecipitation (IP) to enrich for m⁶A-containing mRNA. Of note, the IP step is not required for quantification of m6A stoichiometry at predefined sites, a key utilization of MAZTER-seq. However, applying MAZTER-seq to mRNA that had been subjected to m⁶A-IP step is beneficial for the purpose of *de novo* detection of m⁶A sites and for quality control (QC)ing the performance of MAZTER-seq.

We then examined m⁶A levels at 199 quantifiable ACA sites from the catalog of previously identified putative m⁶A sites that had been computationally inferred by searching for the nearest methylation consensus motif in the proximity of m6A-seq peaks (“m⁶A-seq sites”) (Schwartz et al., 2013). Strikingly, cleavage efficiency scores at m⁶A-seq sites were strongly reduced in the WT samples compared to the ime4Δ/Δ samples, consistent with the presence of methylation at these sites. Moreover, applying m⁶A-IP prior to MAZTER-seq resulted in even greater reduction of the cleavage efficiencies in the WT samples but did not impact their counterparts in the ime4Δ/Δ samples (Figures 2A and 2B). These results thus demonstrate the ability of MAZTER-seq to orthogonally validate putative m⁶A sites at single-nucleotide resolution.

We next sought to detect m⁶A sites *de novo* using MAZTER-seq. We developed an approach relying on three comparisons of cleavage efficiencies: (1) between WT and ime4Δ/Δ Input samples, (2) between WT and ime4Δ/Δ m⁶A-IP samples, and (3) between m⁶A-IP and Input (non-m⁶A-IP) WT samples. For all three comparisons, a true m⁶A site is expected to have lower cleavage efficiencies in the former condition than in the latter. We assembled a database of 45,845 quantifiable ACA sites across the yeast transcriptome, for which we also had sufficient coverage under the surveyed conditions (STAR Methods). We then classified each of these sites into confidence groups, integrating the number of comparisons in which they scored significantly together with the effect sizes and the p values of these comparisons (Figure S3A; STAR Methods), whereby confidence group 0 is the lowest confidence group and 4 is the highest. 410 sites were detected in confidence groups 2 and above, out of which only 56 are part of the previous catalog of m⁶A-seq sites. The following lines of evidence were used to assess the validity of the *de novo* detected sites (“MAZTER-seq sites”), and to conclude that sites detected in the higher confidence groups (two and above) are particularly highly enriched for bona fide m⁶A sites. (1) We estimated a minimal bound on the empirical false detection rate for each of the confidence groups on the basis of the number of significant hits in the reverse comparisons (i.e., sites with enhanced cleavage in the WT strain compared to deletion or in input sample compared to IP) (Figure S3A). While confidence group 1 (i.e., sites scoring significantly only in one comparison) was associated with a substantial false detection rate, confidence groups 2, 3, and 4+ were associated with minimal false detection rates of ~20%, ~7%, and 0%, respectively (Figure 2C). (2) Higher-confidence groups were increasingly enriched in m⁶A-seq sites that had been identified on the basis of peak enrichment (Figure 2D). (3) For each ACA-harboring site in the genome, an “m⁶A-seq score” was calculated on the basis of

m⁶A-seq data from Schwartz et al. (2013), which quantifies the enrichment in coverage at a site in the IP sample in comparison to the Input sample. We found that sites in increasingly higher confidence groups also showed higher m⁶A-seq score levels, providing strong orthogonal evidence for the validity of the sites (Figure 2E). Importantly, such enrichment was clearly evident also when performing this analysis only on sites exclusively identified via MAZTER-seq and not by m⁶A-seq (Figure S3B). (4) Sites in high-confidence groups harbored the same sequence motifs (Figure 2F) and increasing enrichment toward the transcript 3' end (Figure 2G), as was previously reported for m⁶A-seq sites in yeast (Schwartz et al., 2013). (5) m⁶A-seq and MAZTER-seq sites showed similar temporal dynamics across a meiosis time-course (see Figure 5). (6) To directly validate the presence of m⁶A at sites identified by MAZTER-seq, we applied SCARLET, a cleavage- and ligation-based approach interrogating m⁶A levels at individual sites directly by radiolabeling and thin-layer chromatography (Liu et al., 2013), a low-throughput method currently serving as the gold standard in the field. We were able to obtain informative readouts (see STAR Methods) for 14 sites that had been exclusively identified via MAZTER-seq. These included two sites from each of confidence groups 1 and 2 and 5 and 3 sites from confidence groups 3 and 4, respectively. We were able to validate the presence of m⁶A at levels ranging from 7% to 69% at 12 of these 14 sites (Figure 2H). At the two remaining sites, the observed m⁶A signal was not appreciably different from background, indicating either no or very low methylation (for further analyses, the m⁶A level at these sites was assumed to be 0%). Collectively, these analyses provide multiple levels of orthogonal support to the newly detected sites in this collection and demonstrate that the false detection rate, particularly in confidence groups 2–4, is low.

MAZTER-MINE allows us, for the first time, to use an orthogonal approach to estimate the sensitivity and false detection rate of m⁶A sites using m⁶A-IP. The fact that, even in the highest confidence groups, ~4-fold more novel than known sites are detected using MazF based cleavage demonstrates that the antibody-based approach—in their combined experimental and computational implementation—had dramatically underestimated the number of methylation sites. Importantly, having roughly an order of magnitude more m⁶A sites than previously estimated reconciles a previously existing gap between the relatively high levels of m⁶A based on mass spectrometry and the relatively low number of m⁶A sites that had been detected by m⁶A-seq (see STAR Methods).

We also estimated the false-detection rate of antibody-based approaches on the basis of sites that had been detected by m⁶A-seq but were binned into low-confidence groups in MAZTER-seq. Of note, such sites comprise both sites that are not methylated (true negatives) in addition to sites that are methylated but for which we lack the statistical power to assign them into higher-confidence groups. Nonetheless, we observed that sites that had been called by m⁶A-seq but had been assigned into low-confidence groups tended to be substantially more distant from an m⁶A consensus site than their counterparts in higher-confidence groups (Figure S3C). We were moreover able to verify in 4 out of 4 cases that previously called sites—at a large

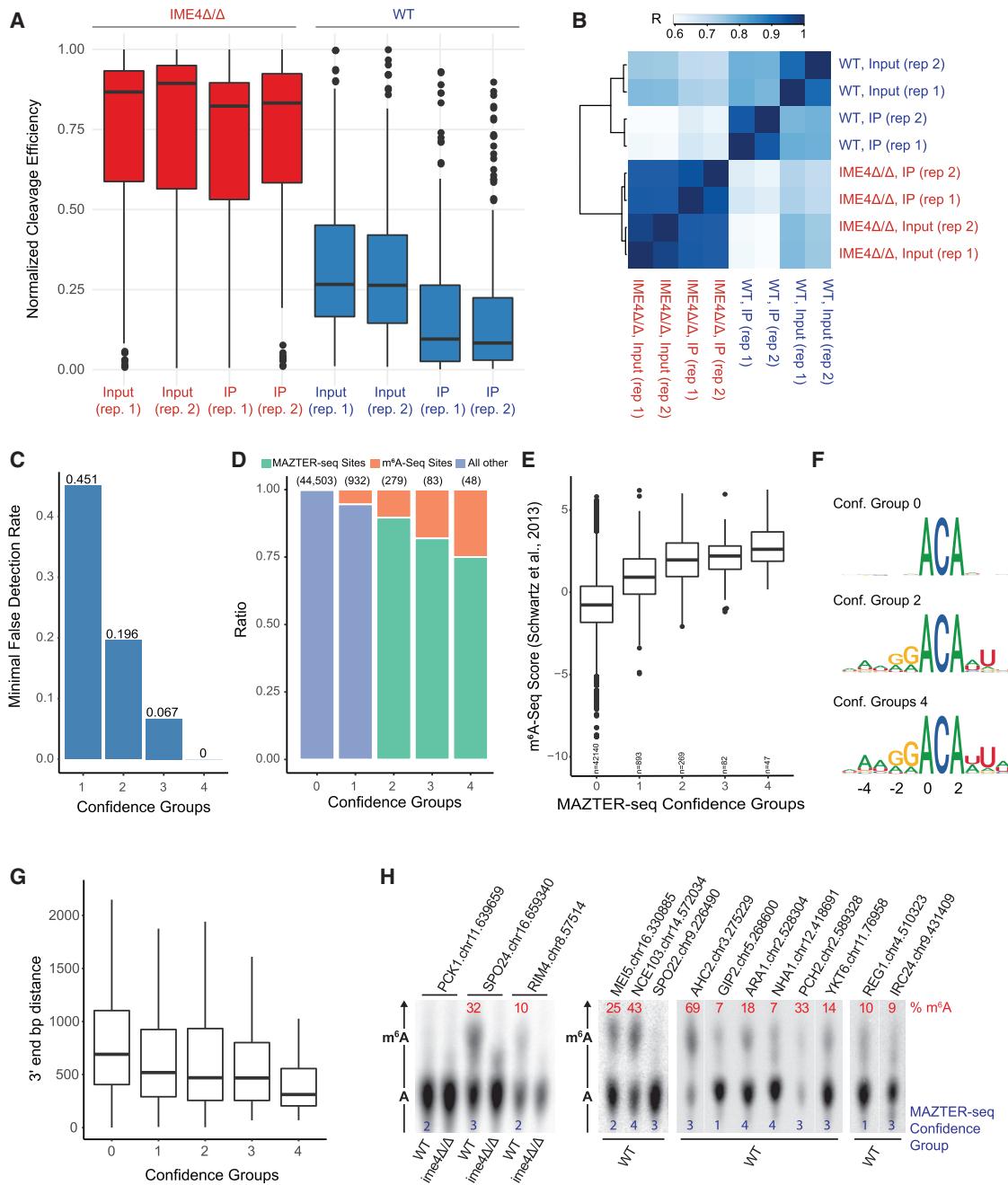


Figure 2. MAZTER-Seq Allows Orthogonal Validation and Discovery of m⁶A Sites

(A) Distribution of cleavage efficiencies (y axis) in RNA extracted from WT *ime4Δ/Δ* strains with versus without m⁶A-IP treatment.

(B) Clustered pairwise correlation of the samples in (A).

(C) Empirical false-detection rates per confidence group.

(D) Distribution of m⁶A-seq sites across the confidence groups defined via MAZTER-seq.

(E) Distribution of m⁶A-seq scores from Schwartz et al. (2013) by MAZTER-seq confidence groups.

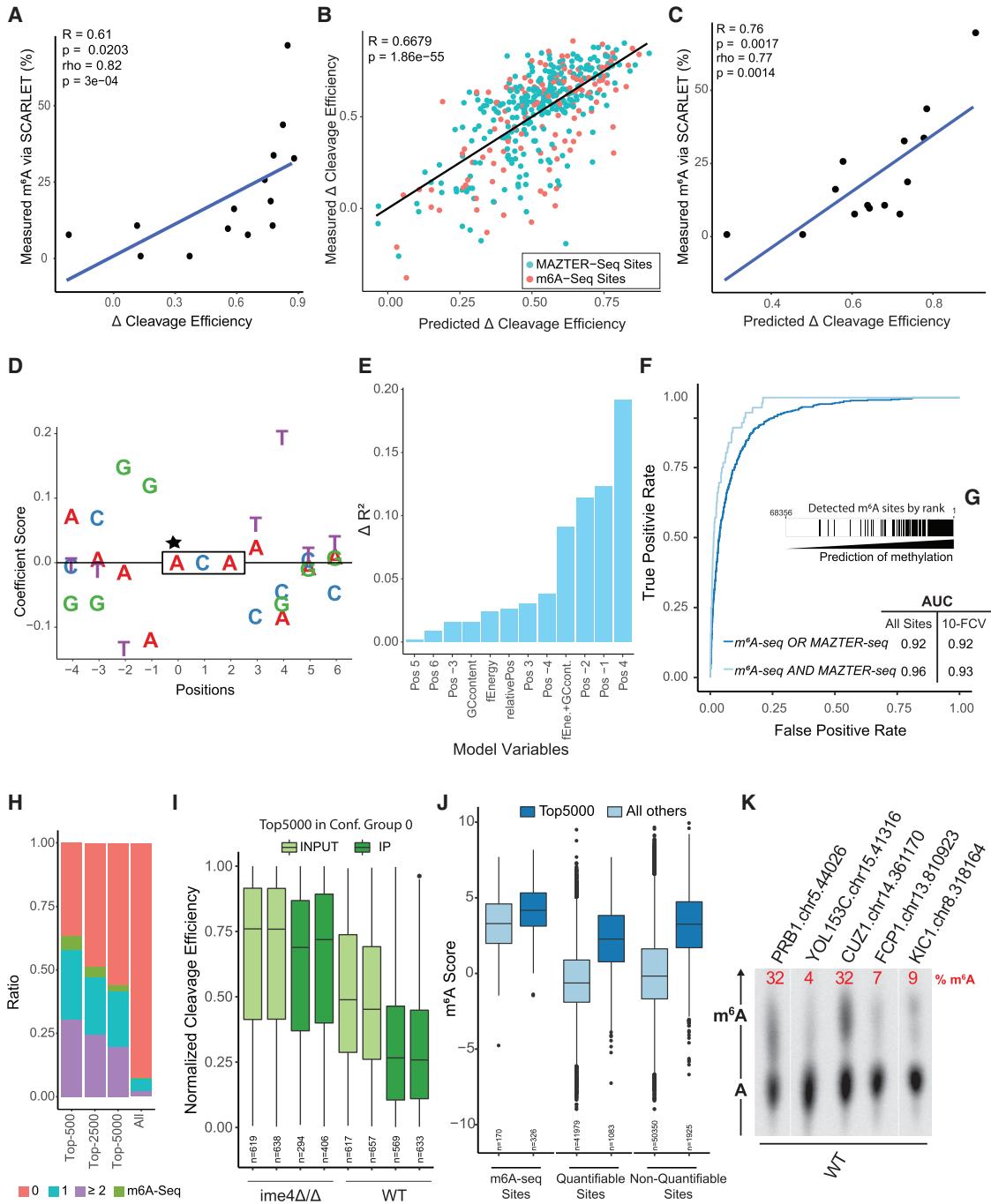
(F) Sequence logos for sites identified via MAZTER-seq shown separately for the indicated confidence groups.

(G) Higher-confidence sites are closer to the end of the transcript. Distributions of 3' end distances by confidence group.

(H) SCARLET-based readouts of methylation levels at each of the indicated sites.

distance from a methylated site—were indeed not methylated (Figure S3D). This analysis allowed us to conservatively estimate a minimal false-detection rate of 11.3% (148 out of 1,308 sites).

MAZTER-seq thus allows us to establish that m⁶A-IP-based approaches are substantially limited in their sensitivity and also suffer from non-negligible false-detection rates.

**Figure 3. m^6A Presence and Levels are Hard-Coded in *cis***

- (A) Correlation between the Δ cleavage efficiency (x axis) and SCARLET quantitations (y axis) for 14 sites.
- (B) Variability of methylation levels can be predicted via local sequence and secondary structure information in both new and previously detected sites. Predicted methylation stoichiometry, via a linear model, is plotted against Δ cleavage efficiency. A regression line fitting all the values is shown in black.
- (C) Correlation between the linear-model predicted cleavage efficiency (x axis) and the SCARLET quantitations (y axis) for 14 sites.
- (D) Plot depicting the relative contribution to methylation of different nucleotides surrounding the m^6A site. x axis shows relative position with respect to the m^6A site; y axis shows model-derived coefficients value for each nucleotide at a given position. To facilitate interpretation, each site's coefficients were first scaled to a mean of 0.
- (E) Relative contribution of each of the indicated variables to the performance of the stoichiometry model. Depicted is the difference in R^2 from the full model when removing each of the variables by one-in-one-out.

(legend continued on next page)

m⁶A Deposition and Stoichiometry Are Hard Coded into the Yeast Genome

We next sought to explore the extent to which methylation stoichiometries can be predicted. Toward this goal, we first sought to identify the optimal quantitative measure of m⁶A levels. If MazF cleavage were 100% efficient at unmethylated sites, all ACA sites in an ime4Δ/Δ strain (which lacks m⁶A) should be cut at 100% efficiency. However, inspection of the distribution of this metric revealed that cutting by MazF is variable from one site to another (Figure 2A). Two factors contributed to these effects in a predictable manner: (1) sequence composition immediately flanking the ACA site that predictably impacted MazF cleavage (Figures S4A–S4C), and (2) predicted single strandedness of the region surrounding the ACA site, which correlated significantly with cleavage scores ($R = 0.3$, permutation test [10,000 permutations] $p < 1 \times 10^{-3}$), consistent with previous observations that MazF is biased toward cleavage of single-stranded RNA (Zhang et al., 2003) (Figure S4D). As a quantitative metric of RNA methylation, we thus introduced a Δcleavage-efficiency metric, capturing the difference between MazF cutting in WT versus ime4Δ/Δ strains, reasoning that this difference will eliminate effects originating from distinct baseline cutting levels. Indeed, Δcleavage-efficiency lost most of its previously existing association with both sequence and secondary structure (Figure S4E) and with sequence (Figure S4F).

The quantitative performance of Δcleavage-efficiency is supported by several critical lines of evidence. First, and most importantly, Δcleavage-efficiency levels correlate highly with SCARLET-based quantitations, which serve as a gold standard for m⁶A quantitation (Spearman's rho = 0.79, $p = 1 \times 10^{-3}$; Figure 3A). Raw cleavage efficiencies also correlated highly, but slightly more poorly, with SCARLET quantitations (rho = 0.78, $p = 1 \times 10^{-3}$). Second, we observed a significant correlation between Δcleavage-efficiency and the m⁶A-seq scores ($R = 0.38$, permutation test [10,000 permutations] $p < 1 \times 10^{-3}$), which are derived based on orthogonal measurements. Raw cleavage efficiencies correlated slightly more poorly with m⁶A-seq scores ($R = -0.37$, permutation test [10,000 permutations] $p < 1 \times 10^{-3}$). Δcleavage-efficiency also correlated better with the fold change in enrichment in coverage upon IP in the WT strain with respect to the coverage upon IP in the ime4Δ/Δ strain ($R = 0.44$ and $R = -0.41$, respectively). These data suggest that while both metrics perform reasonably well in capturing methylation stoichiometry, Δcleavage-efficiency provides a slightly more reliable relative measurement of methylation levels.

We next trained a simple linear model using the base identity of 4 bp upstream (positions –1 to –4) and 4 bp downstream (positions 3 to 6) of the ACA mazF cutting consensus sequence (note that positions 0, 1, and 2 are fixed) alongside three features capturing the GC-content, the propensity toward secondary structure (predicted “free energy”), and the relative position within the gene. We trained this model based on sites identified exclusively via MAZTER-seq and used the sites identified via m⁶A-seq as a validation set. The linear model yielded excellent agreement with the MAZTER-seq-derived Δcleavage-efficiencies both in the training set ($R^2 = 0.46$) and in the validation set ($R^2 = 0.37$) (Figure S5A). A model based on both MAZTER-seq and m⁶A-seq sites achieved an $R^2 = 0.45$ (Figure 3B). As an independent validation of the model, we applied the model to the set of sites measured via SCARLET and obtained an excellent agreement between the predictions of the model and the measured values ($R = 0.78$, $p = 1 \times 10^{-3}$; Figure 3C). Critically, training an identical model on the same sites but using m⁶A-seq scores (instead of Δcleavage-efficiencies) explained much less of the variance ($R^2 = 0.11$) when applied to the same training-validation scheme, reflecting the reduced quantitative nature of m⁶A-seq (Figure S5B). These findings both establish the predictive power of the model and further support the validity of the novel set of sites. Our results thus strongly suggest that in yeast, m⁶A stoichiometries are primarily dictated in *cis* via a simple code embedded in the sequence and structure at the modified site.

To identify and rank the factors contributing to the performance of the model, we examined the coefficients assigned to them by fitting the model (Figure 3D; Table S1) and assessed the relative contribution of the variables in the model by removing the variables in a one-in-one-out fashion and calculating the difference in the resulting R^2 from the original R^2 (ΔR^2) (Figure 3E). The variables contributing most to the model included not only positions –1 and –2, both of which favored a “G,” consistent with Dominissini et al. (2012) and Meyer et al. (2012), but also positions +4 and –4 with respect to the m⁶A site, with positive effects on stoichiometry for an A at the former and a “U” at the latter, consistent with Schwartz et al. (2013). Other features, mostly with reduced contribution, included lack of secondary structure at the region surrounding the modified site along with GC content, also consistent with previous observations, and relative position within the gene, with a bias toward 3' end (Schwartz et al., 2013). These results thus suggest that nearly 50% of the variability in m⁶A levels from one site to another is determined primarily via local sequence, with minor

(F) Receiver operator curves (ROCs) of two predictors of methylation deposition for sequences with the RRACA consensus motif. Both predictors were also tested using 10-fold cross-validation, and the resulting AUCs are showed.

(G) Visual depiction of the model predictions and already-detected sites. All RRACA consensus sites are sorted based on their predicted score; a black bar (above) denotes whether methylation was identified at the site (either using m⁶A-seq or MAZTER-seq).

(H) For each top-n model predictions ($n = 500, 2,500, 5,000$), the distribution of confidence groups among these sets of sites are depicted (note: only for the subset of quantifiable sites). In addition, if the site was identified via m⁶A-seq (but not MAZTER-seq), this is indicated in a separate color. All sites are shown for comparison, serving as background.

(I) Quantifiable top 5,000 scoring sites show decreased levels of cleavage in the WT compared with the deletion strain and further decrease upon m⁶A-IP. Distributions of cleavage efficiency (y axis) across WT and ime4Δ/Δ yeast cells, either with or without m⁶A-IP treatment, are shown

(J) Top 5,000 sites show extensive enrichment upon m⁶A-IP in comparison to all remaining sites. Results are shown separately for quantifiable and non-quantifiable sites (via MAZTER-seq). As a control, distributions are also shown for sites identified via m⁶A-seq.

(K) SCARLET readouts for five sites that were non-quantifiable via MAZTER-seq and were predicted purely based on the logistic model.

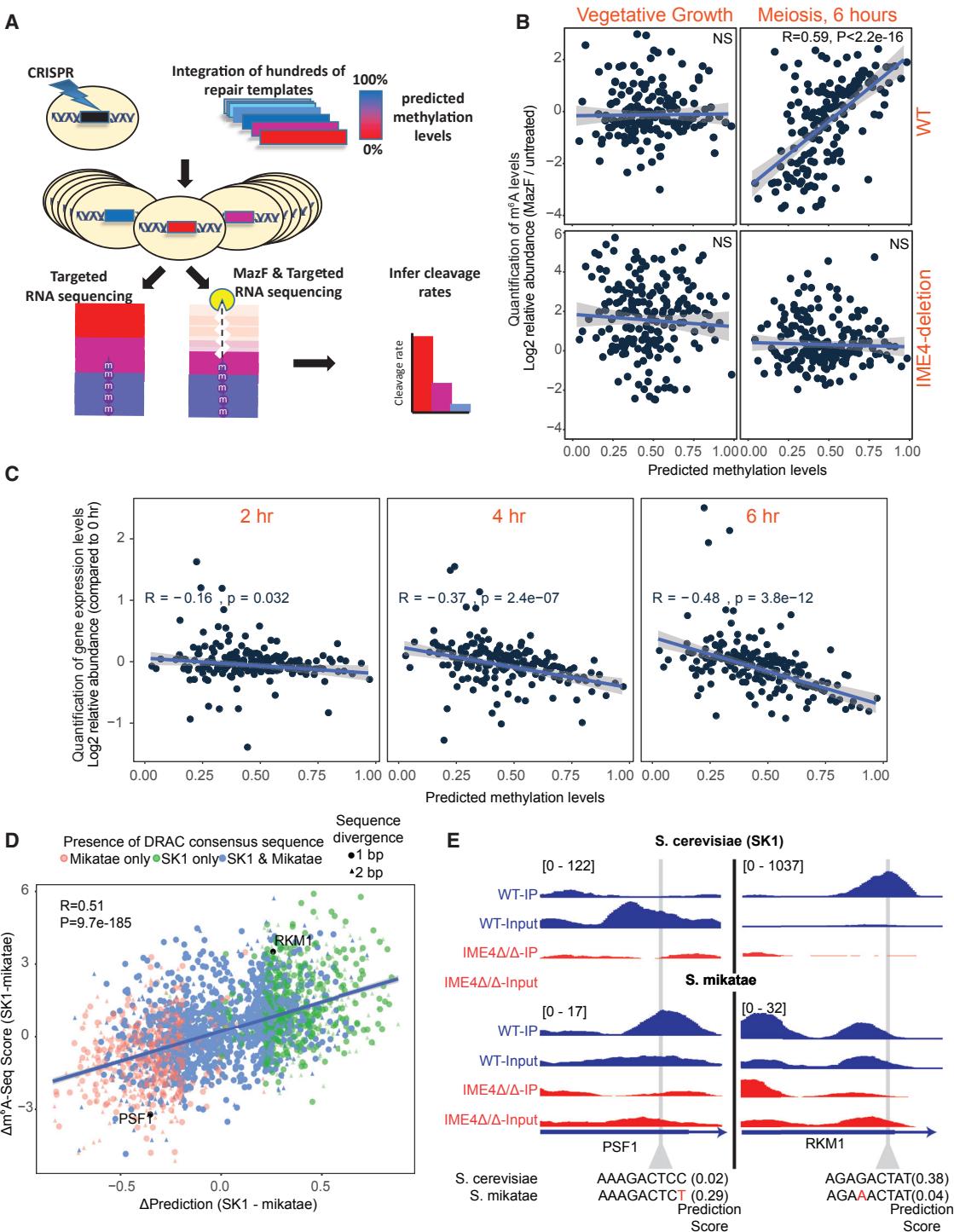


Figure 4. Systematic Validation of Model via Experimentally Induced and Natural Perturbations

(A) Scheme depicting the experimental approach to introduce and measure m⁶A levels at hundreds of variants.

(B) Correlations of predicted methylation levels with cleavage efficiencies.

(C) Scatterplots depicting the correlations between predicted methylation levels and between the relative abundance of each variant at the indicated time points with respect to the 0-h time point.

(D) Correlation between the difference in predicted methylation of orthologous m⁶A consensus motifs between SK1 and *S. mikatae* (Δ prediction, x axis) and the difference in m⁶A-seq scores at the same sites between SK1 and *S. mikatae* (Δ m⁶A-seq score, y axis). Each site is plotted as a circle if the sequence divergence in the 9-bp window centered around the methylation site is 1 bp or a triangle if the divergence is 2 bp. The color of each point reflects whether the core sequence

(legend continued on next page)

contributions from secondary structure and the proximity of the site to the end of the gene.

We next inquired whether m⁶A presence, at a genome-wide scale, was similarly hard coded and predictable. For this task, we generated an ultra-high-confidence set of m⁶A sites, comprising all sites identified both in m⁶A-seq and via MAZTER-seq (defined as sites in confidence groups ≥ 2) and harboring an RRACA consensus motif (R = A/G). Remarkably, a logistic classifier trained with these sites as positives and all remaining transcriptomic RRACA sites that had not been flagged as potential m⁶A sites as negatives was able to achieve an area under the curve (AUC) of 0.96 (0.93 in a 10-fold cross-validation setting), indicating an excellent ability to discriminate between methylated and non-methylated sites (Figures 3F and 3G). A model generated on the basis of sites defined via m⁶A-seq or MAZTER-seq had similar performance (AUC = 0.92) (Figures 3F and 3G). Of note, the relative weights assigned by the linear and by the logistic classifier were overall very similar (Tables S1 and S2). Consequently, the predictions yielded by the two models were highly correlated ($\rho = 0.9$), and the relative contribution of different variables were similar (Figure S5C).

We next sought to assess the ability of this model to detect sites *de novo* at a genome-wide level. We applied the above-derived logistic model to each of the 68,356 RRACA sites in the yeast transcriptome (Table S4) and examined the 5,000 sites with the highest scores (top 5,000). Remarkably, nearly 50% of the quantifiable sites, were classified in confidence groups 1 and upward, a massive enrichment with respect to the background (7.6%) (Figures 3H and S6A). Note that the model was trained on sites in confidence groups ≥ 2 ; thus, the enrichment for confidence group 1 (which is enriched 4-fold more than expected; Figure S6A) cannot be due to overfitting. As indicated above, many of the sites in confidence group 1 and the vast majority of sites in confidence group 0 do not reflect truly methylated sites. However, examination of the cleavage efficiencies in the sites forming part of the top 5,000 showed that even when these sites were assigned to confidence group 0, they showed strong evidence of decreased cleavage in WT compared to deletion (Figure 3I) and a substantial enrichment upon m⁶A-IP compared to Input (Figures 3J and S6B). Moreover, the non-quantifiable sites forming part of the top 5,000—i.e., the set of sites for which we had been unable to obtain any measurements using MAZTER-seq and that had been predicted exclusively based on the model—were also dramatically enriched upon m⁶A-IP, in fact to a similar extent as the m⁶A-seq sites that had originally been detected (Schwartz et al., 2013) (Figure 3I), demonstrating the ability of the model to *de novo* detect m⁶A sites. Finally, the set of top 5,000 sites also showed similar temporal dynamics in a meiosis experiment, as the well established m⁶A-seq sites (see below and Figure S6C). These results thus strongly suggest that a substantial proportion of the top-scoring predicted sites—even the ones corresponding to classes 0 and 1—are likely truly methylated. Their assignment to

lower-confidence groups likely reflects the lack of statistical power to classify them into higher bins, most probably due to their decreased coverage (Figure S5D).

Finally, to orthogonally validate this genome-wide model, we selected five sites from the top 5,000 predictions by the model. Importantly, all of these sites were non-quantifiable via MAZTER-seq (as they lie within very close proximity to both upstream and downstream ACA sites) and do not form part of the m⁶A-seq sites (Schwartz et al., 2013). Remarkably, we were able to validate the presence of methylation at all five sites via SCARLET at stoichiometries ranging from 4% to 32% (Figure 3K), demonstrating the power of this model and the predictable pervasiveness of m⁶A across the yeast transcriptome.

Systematic Validation of Model via Large-Scale Targeted Perturbations

To experimentally test the model, we used CRISPR/Cas9 to generate a library of yeast mutants harboring hundreds of designed sequence variants of the m⁶A consensus site, each associated with a predicted level of methylation. Methylation levels at each of the mutants was measured, as a pool, by extracting mRNA during meiotic prophase and quantifying the abundance of each of the variants before and after cleavage with MazF, with the expectation that increased stoichiometries of methylation at the target site would result in increased protection from MazF cleavage (Figure 4A). The predicted levels by the model agreed well with the measured ones ($R = 0.59$, $p < 2.2 \times 10^{-16}$) in WT cells under meiosis conditions (Figure 4B) and was completely eliminated under vegetative conditions and in *ime4Δ/Δ* strains (Figure 4B). This set of experiments thus directly, and systematically, validates the ability of our model to quantitatively predict m⁶A levels.

The ability to quantitatively perturb m⁶A levels offered a unique opportunity to address questions pertaining to the impact of m⁶A on mRNA fate using a dose-response experimental design. We utilized the above-described pool of yeast strains harboring hundreds of yeast variants at an individual site and quantified the levels of each of the sequence variants across meiosis at 2-h time-point intervals culminating at 6 h, in which methylation levels peak. Remarkably, we observed a constantly increasing, inverse correlation between the predicted methylation levels of each variant and its relative abundance, from $R = -0.16$ (at the 2-h time point) to $R = -0.37$ (4 h) to $R = -0.48$ (6 h) (Figure 4C). This analysis thus strongly suggests that m⁶A negatively regulates mRNA expression levels in yeast. In mammalian systems, comparison of stability in the presence versus absence of METTL3 has shown that methylated genes have an increased half-life upon depletion of METTL3. However, studies have consistently been able to show a very weak—albeit statistically significant—effect for these analyses (Herzog et al., 2017; Wang et al., 2014). Our ability to now quantitatively perturb methylated sites allows to establish substantially stronger dose-response relationships between m⁶A levels and a molecular readout.

motif—defined as presence of a DRAC core (D=A,G,T; R=A,G)—is present only in *S. mikatae* (red), only in *S. cerevisiae* (green), or in both species (blue). The two labeled points indicate the sites plotted in (E).

(E) Visualization of coverage tracks—obtained from m⁶A-IP experiments—in both SK1 and *S. mikatae* for the two sites highlighted in (D) showing sequence changes associated with differential methylation between the two species.

Single Point Mutations Are Sufficient to Drive Predictable Loss and Acquisition of Methylation Sites across Evolution

Our findings that m⁶A is dictated, to such a high extent, in *cis* suggests that changes in sequence—as occur naturally throughout evolution—should give rise to predictable loss and acquisition of methylation sites. To investigate this, we applied the above derived logistic model to all DRAC-containing sites in two yeast species: *S. cerevisiae* and *S. mikatae*. We identified all sites predicted to undergo methylation in either of the species and calculated a Δprediction score capturing the difference in their predicted likelihood of undergoing methylation in *S. cerevisiae* versus *S. mikatae*. We then made use of available m⁶A-seq datasets in meiosis for both organisms (Schwartz et al., 2013) and calculated for each site the Δenrichment score, defined as the difference between the m⁶A-seq score in *S. cerevisiae* and *S. mikatae*. To simplify the interpretation of this comparison, we limited our analyses to cases where the 9-bp methylation consensus window diverged by at most two base pairs between the two species. We observed a striking positive correlation between the Δprediction and Δenrichment scores ($R = 0.51$, $p = 9.7 \times 10^{-85}$; Figures 4D and 4E), indicating that loss or acquisition of a methylation site occur, in a predictable manner, through single point mutations occurring across evolution.

Quantitative Tracking of m⁶A Levels across Yeast Meiosis

The inability to quantify m⁶A levels has been particularly limiting in the context of dynamic cellular- and/or disease-related responses. To evaluate the ability of MAZTER-seq to quantify m⁶A levels across a dynamic response, we applied it to a densely profiled time course following induction of meiosis in yeast (Figure 5A). We observed a gradual increase in methylation levels up to the 6-h time point, which coincides roughly with meiotic prophase, followed by a reduction in the subsequent time points (Figure 5B). The median methylation levels derived from MAZTER-seq were in strong agreement with mass-spectrometry-based measurements of m⁶A levels ($R = -0.8$) (Figure 5C). Remarkably, the measurable set of top 5,000 predicted sites displayed highly similar dynamics to those displayed by m⁶A-seq and MAZTER-seq sites, lending strong additional support to the validity of the model for *de novo* prediction of methylation levels. Finally, these results further suggest that while the methylation potential of each site may be encoded in *cis*, leading to low potential for local regulation of m⁶A levels, global regulation of m⁶A can be achieved likely through titration of the concentrations of the methyltransferase machinery.

MazF-qPCR Allows Rapid Readouts on m⁶A Levels at Individual Loci

One limitation of m⁶A-seq—and particularly its cross-linking-based derivatives—was that it typically had to be tied to a sequencing readout and in our hands yielded variable results when coupled with qPCR. This, in conjunction with the high amounts of material required for m⁶A-IP, has limited the ability to design a cheap and robust quantitative readout of m⁶A levels to be used for example in the context of genetic screens. To test

whether MazF-based digestion could provide such a readout, we designed MazF-qPCR, a qPCR-based assay with which we targeted methylation sites in three yeast genes (Figures S7A and S7B; STAR Methods). We could confirm reduced cleavage levels were observed in WT strains in prophase in comparison to *ime4Δ/Δ* (Figure 5D). Applying MazF-qPCR across a full meiotic time course, we could moreover recapitulate the gradual increase in m⁶A levels up to prophase, followed by its subsequent decrease (Figures 5E and S7A). MazF-qPCR was also successfully applied to target two m⁶A sites in mouse (Figure S7B). MazF-qPCR is thus amenable for cheap and robust interrogation of m⁶A levels using 25–50 ng of mRNA as input amounts.

m⁶A Code Is Conserved in Mammalian Systems

We next evaluated the applicability of MAZTER-seq in mammals and applied MAZTER-seq to WT and METTL3 knockout (KO) mouse embryonic stem cells (mESCs). To assess the performance of MAZTER-seq, we assembled a catalog of ACA sites in mouse divided into low-, intermediate-, and high-confidence sites based on available m⁶A-seq data in mouse tissues (Schwartz et al., 2014a); in addition, we assigned sites as “highest-confidence” if they had also been detected as methylated using m⁶A individual-nucleotide-resolution cross-linking and immunoprecipitation (miCLIP) (Linder et al., 2015) (Table S5). MAZTER-seq measurements confirmed that cleavage efficiencies progressively decreased in higher-confidence groups in WT cells. In contrast, upon METTL3 KO, the distributions of cleavage efficiencies across the different confidence groups were roughly identical (Figure 6A; Table S5).

We further found that a simple code—primarily capturing the sequence at the modified site—was able to capture 35% of the variability in Δcleavage-efficiency (Figure 6B), suggesting that also in mammals, a major portion of the m⁶A signal is hard coded, though possibly to a reduced level than in yeast (see Discussion). Remarkably, the features contributing to this code were qualitatively reminiscent of the ones identified in yeast, including the bias toward an A at position –4 and a “T” at position +4, along with the biases toward a G at positions –1 and –2. Such an extended consensus sequence has not previously been identified in mammals (Figures 6C and 6D). Using an analogous approach to the one described above for yeast, we could also *de novo* detect m⁶A sites. As in yeast, higher-confidence bins were progressively enriched with previously identified sites (Figure 6E), yet 70% of the sites in the highest-confidence bin had not been previously detected with m⁶A-seq, suggesting that the sensitivity of detection via m⁶A-seq had been low (Figure 6E). The newly identified sites could be systematically validated via m⁶A-seq and were enriched by m⁶A-seq to similar levels as sites that had been called based on this technique (Figure 6F). The principles of the m⁶A code—and most likely also their mechanistic underpinnings—are thus to a large extent conserved between yeast and mouse.

MAZTER-Seq Allows Quantitative Evaluation of Sites Identified via miCLIP

The current state of the art approach for global identification m⁶A sites at single-nucleotide resolution is miCLIP (Linder et al., 2015). However, the overlap between miCLIP sites identified in

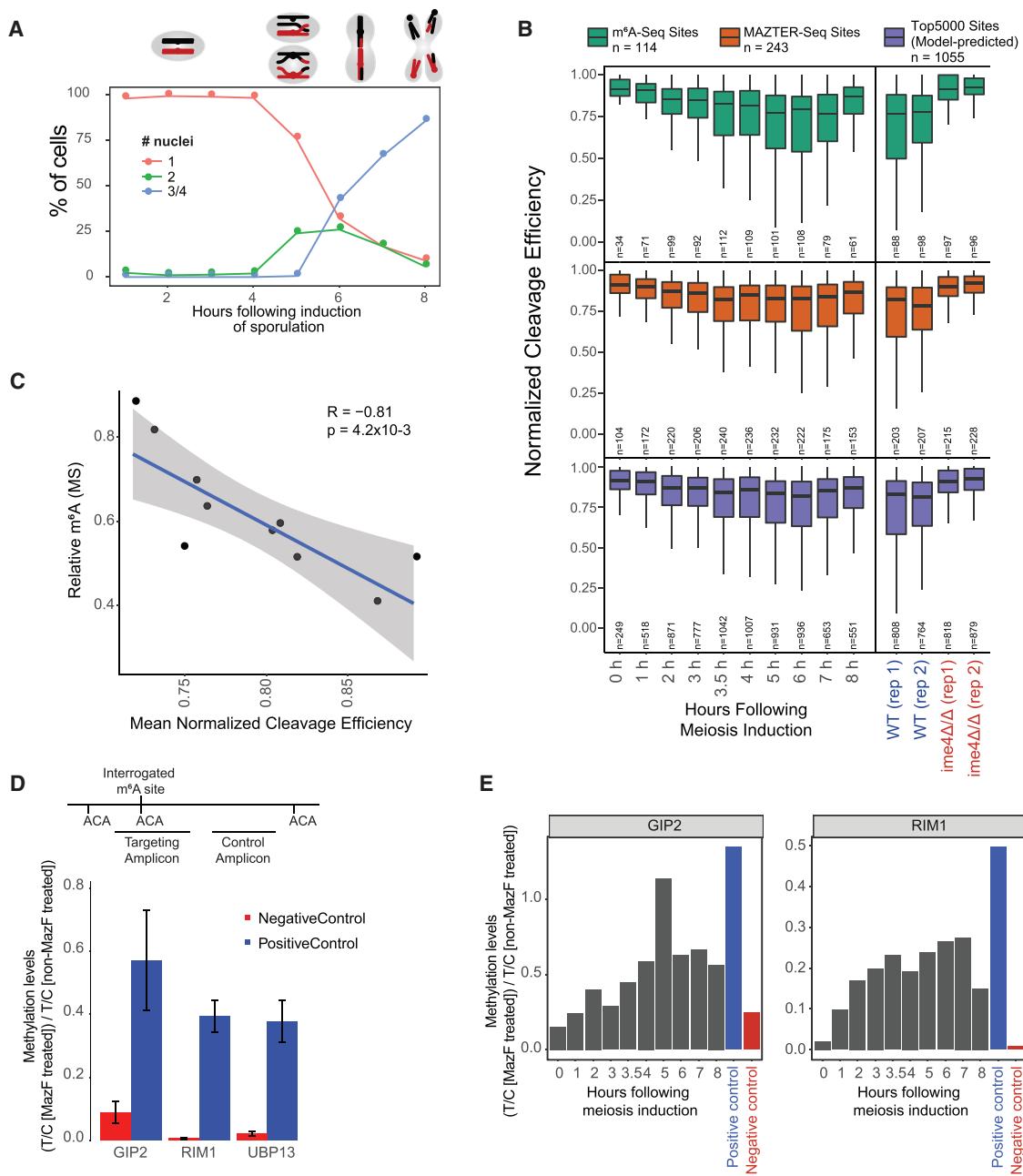
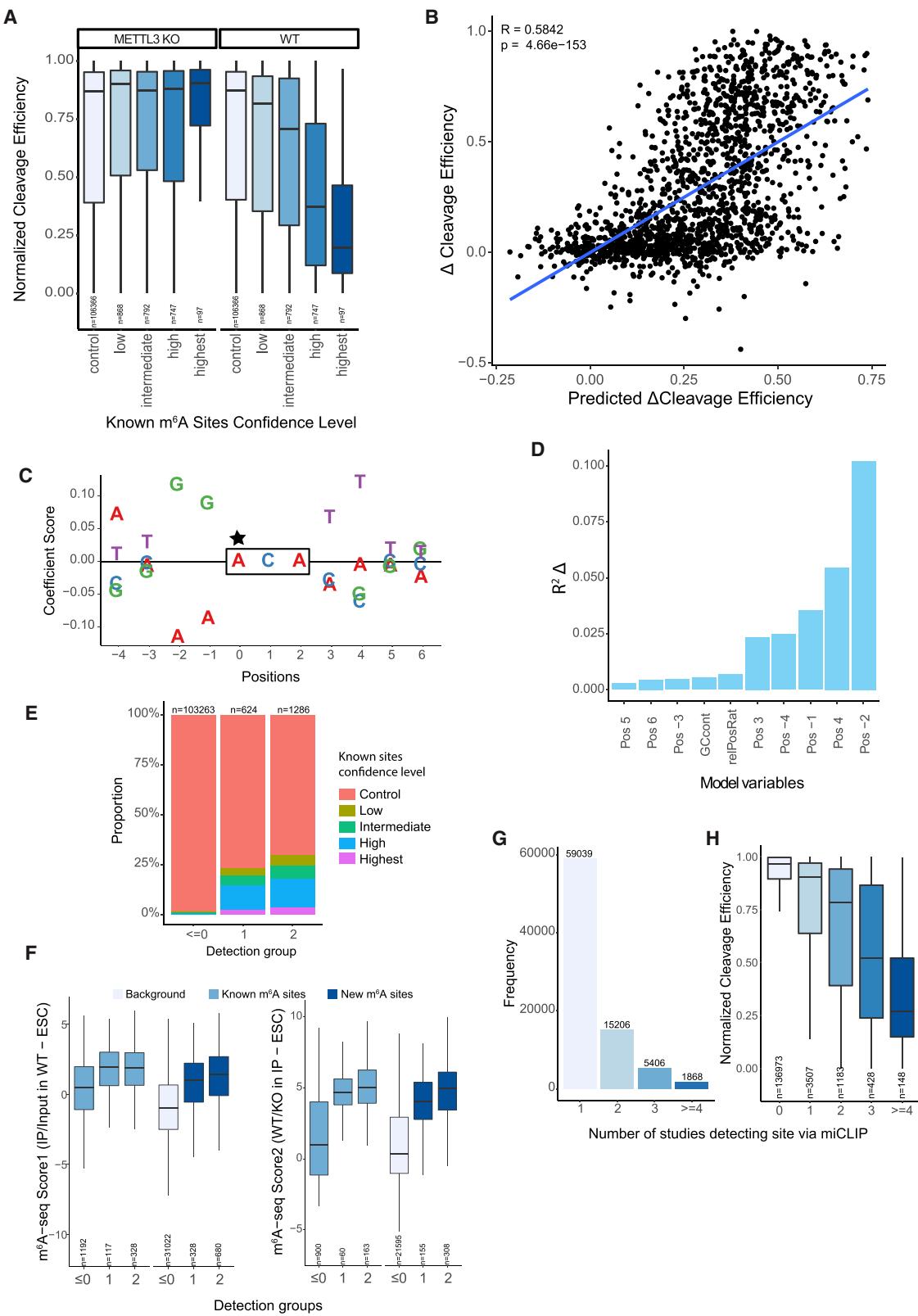


Figure 5. m⁶A Dynamics in Meiosis

- (A) Percentage of single, double, triple, or tetra nuclei cells counted for the different time points based on DAPI staining at indicated time points.
- (B) Meiotic m⁶A dynamics measured with MAZTER-seq. Distributions of normalized cleavage efficiencies for three different groups of m⁶A sites: m⁶A-seq sites, MAZTER-seq sites, and model-predicted sites.
- (C) Correlation between mass-spectrometry-based quantitations of m⁶A levels and mean normalized cleavage efficiencies derived from MAZTER-MINE.
- (D) Bars represent the methylation levels measured via MazF-qPCR, a qPCR-based assay coupled with mazF digestion, for a 10-point meiosis time-course. The levels of a targeted amplicon (labeled "T") is measured against a control (labeled "C") amplicon in a mazF-digested sample and normalized against a non-digested sample.
- (E) MazF-qPCR across a meiotic time course. Positive and negative controls (strains SAy841 and SAy966, 8 h following induction of meiosis) are also reported.

different studies is typically quite low. For example, examination of 81,519 sites, identified across six different published miCLIP experiments in human cell lines, reveals that the vast majority of these sites are identified only in a single study (Figure 6G),

with only ~1,500 sites, on median, being shared between any two datasets. While this could be due to technical noise, or to biological variability in m⁶A topologies between samples, sites that are not reproducibly detected could also reflect sites



(legend on next page)

modified at low stoichiometries that are stochastically identified in one miCLIP experiment, but not in another, due to the low efficiency of the UV crosslinking. Binning m^6A sites based on the number of miCLIP experiments in which they are identified (Figure 6H), we found that (1) even sites detected in only a single study are generally cleaved at lower efficiencies than controls, suggesting that many of these sites are indeed methylated, and (2) the number of miCLIP experiments in which an m^6A site is detected strongly correlates with the stoichiometry of m^6A , as estimated by MAZTER-seq. Our results thus strongly suggest that any individual miCLIP experiment provides a stochastic sampling of a subset of modified sites and that by combining MAZTER-seq measurements with miCLIP experiments—and integrating multiple miCLIP experiments—it becomes possible to focus on high-stoichiometry sites, which are more likely to mediate function. These findings further suggest that the majority of the ~80,000 sites identified across any of the samples are likely truly modified, consistent with our inferences both in yeast and in mouse regarding the limited sensitivity of antibody-based approaches.

Quantitative Profiling of m^6A in Development and upon Subtle Perturbations of Mammalian Systems

MAZTER-seq has the potential to resolve fundamental questions pertaining to the extent to which m^6A can be dynamically modulated through m^6A “erasers.” Specifically, the discovery of the m^6A demethylase fat-mass- and obesity-associated protein (FTO) (Jia et al., 2011)—and later of ALKBH5 (Zheng et al., 2013)—prompted the hypothesis that m^6A is dynamically modulated within cells, or across conditions, through active demethylation. Yet to date, the support for strongly regulated m^6A levels across subcellular compartments (Chen-Kiang et al., 1979; Nevin and Darnell, 1978) or between conditions (Schwartz et al., 2014b) is quite limited, albeit with the caveat that the methodologies used in these analysis (m^6A -seq or chromatography) are limited in their ability to quantify such changes. Indeed, recent studies, again based on bulk analysis of m^6A content, have reported evidence for selective nuclear demethylation of m^6A by FTO (Wei et al., 2018). We therefore examined m^6A levels upon either KO or overexpression of FTO in human embryonic stem cells (hESCs) and HEK293T cells, respectively (Figures S7C and S7D; Table S6). In both cases, we did not observe an impact

on m^6A levels (Figures 7A–7D). In contrast, as a positive control, when we overexpressed the m^6A demethylase ALKBH5 (Zheng et al., 2013), we found a subtle but reproducible decrease in methylation levels, which became increasingly pronounced at higher-confidence sites (Figures 7E and 7F). Our results are thus consistent with recent reports suggesting that rather than demethylating m^6A on mRNA, FTO primarily demethylates m^6Am (a related but distinct modification) on small nuclear RNA (snRNA) (Mauer et al., 2017, 2019).

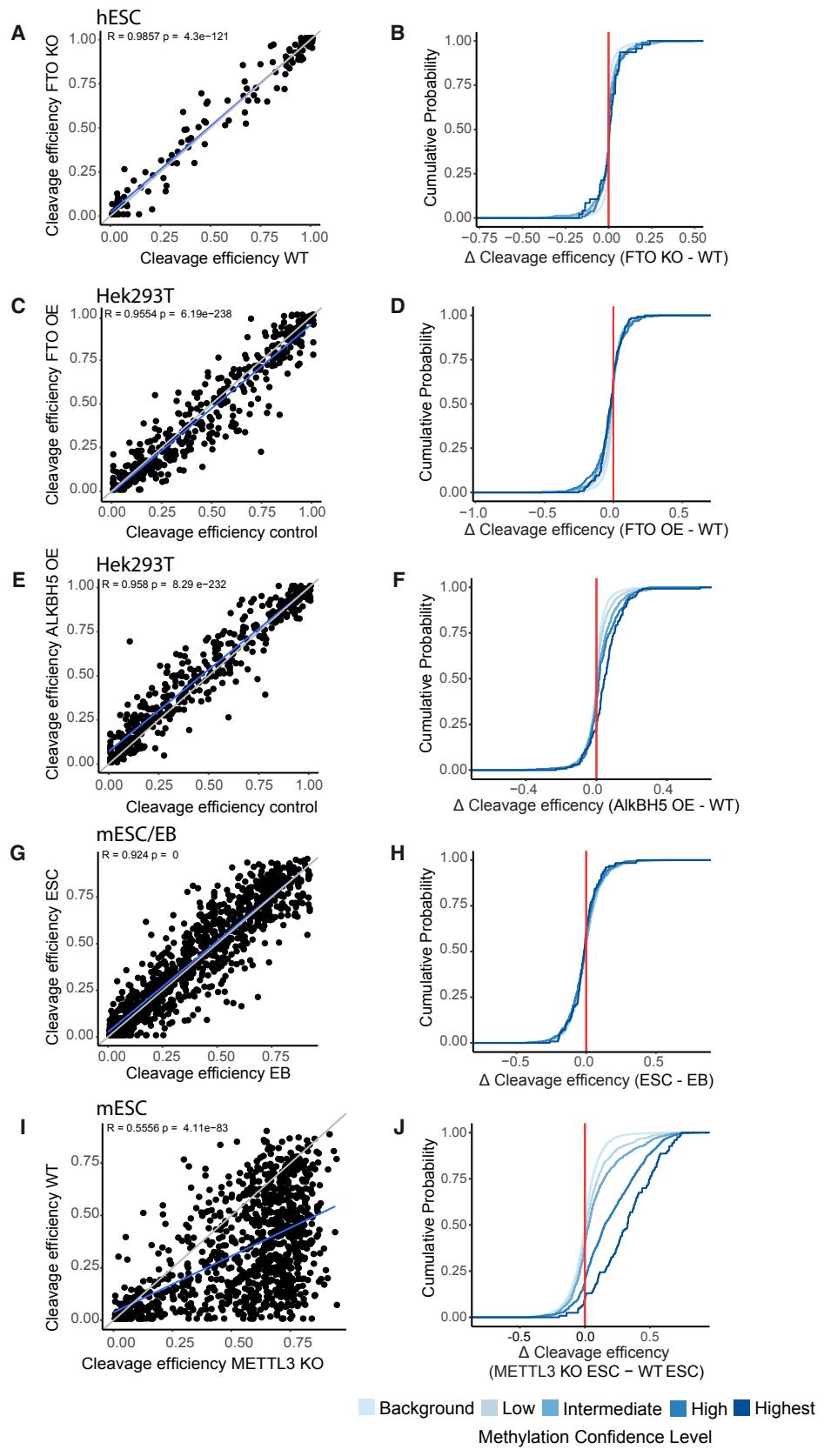
Finally, we chose to compare m^6A levels in mouse mESCs to their counterparts in embryoid bodies (EBs) into which they were differentiated based on previous findings that knocking out METTL3 in mESCs leads to a differentiation defect that is apparent already within EBs (Geula et al., 2015). Accordingly, we differentiated either WT or METTL3-KO mESCs into EBs and applied MAZTER-seq to both cell types under both backgrounds. We observed that cleavage efficiencies across the modified sites (excluding low-confidence sites) were highly correlated between ESCs and EBs, demonstrating that no global scaling of m^6A levels occurs at the transition from ESCs to EBs (Figures 7G and 7H). In contrast, comparison of WT to METTL3 KO mESCs revealed a dramatic decrease in methylation (Figures 7I and 7J). These findings demonstrate that the requirement of the methyltransferase complex in the transition from ESCs to EBs probably does not reflect a global redistribution of m^6A ; instead, it could potentially reflect a change in how m^6A is “interpreted”—for instance, by diverse reader proteins.

DISCUSSION

MAZTER-seq provides an orthogonal quantitative readout of methylation levels and now allows us, among others, to dissect the m^6A code, to explore its function, and to track methylation methylation levels across dynamic responses and genetic perturbations. A key finding in this study is the remarkable simplicity of the m^6A code, accounting for 33%–46% of the variance in m^6A levels. The simplicity of this code suggests that despite the presence of diverse m^6A readers and erasers, their ability to act in a local, site-specific manner is likely quite limited. Instead, the primary mode through which m^6A is likely to undergo modulation is through global regulation of m^6A levels, as observed across the meiotic time course (Figure 4C). Such

Figure 6. m^6A Code Is Conserved in Mammals

- (A) MAZTER-seq quantification of cleavage efficiencies at mouse m^6A sites. Sites are divided into four confidence groups based on antibody-based measurements in addition to a control group in which methylations had not been detected (STAR Methods).
- (B) Correlation between Δ cleavage efficiencies (y axis) and the predicted Δ cleavage efficiencies by the mouse linear model.
- (C) Depiction of coefficients assigned by the linear model to each of the nucleotides 4 nt upstream and 6 nt downstream of the m^6A site. To facilitate interpretation, the coefficients were scaled to a mean of 0.
- (D) Contribution of variables to the performance of the model measured by difference in R^2 .
- (E) Proportion of previously detected sites (colored based on their confidence level) across three detection groups defined on the extent of reduced cleavage in WT in comparison to METTL3-KO mESCs.
- (F) m^6A -seq-based validation of newly detected sites by MAZTER-seq. Enrichment in m^6A -seq for each of the three confidence bins are shown for both previously detected sites using antibody-based methods and for novel sites detected exclusively via MAZTER-seq. Two metrics of enrichment by m^6A -seq are shown: enrichment in IP compared to Input (left) and enrichment in IP in a WT background in comparison to an METTL3 KO mutant (right).
- (G) Analysis of overlap between different miCLIP datasets. Number of miCLIP experiments (out of a maximum of 6) in which a site was detected is shown on the x axis; number of sites is shown on the y axis.
- (H) MAZTER-seq-based quantitations of cleavage efficiencies at sites identified in miCLIP and binned based on the number of miCLIP experiments in which they had originally been identified. MAZTER-seq measurements were performed in cytosolically enriched RNA from HEK293T cells.



regulation is most likely achieved through titration of the levels of the methyltransferase complex components, leading to a global scaling of methylation levels, as is the case in meiosis (see Figure S7E). Nonetheless, it is possible that the somewhat reduced proportion of the variability explained by our model in mESCs compared to yeast could reflect the presence of only a single reader and absence of any reported demethylases in yeast in contrast to the more complex landscape of readers and erasers in mammals.

MAZTER-seq also suffers from limitations. (1) It allows quantification of only a subset of m⁶A sites that both occur at ACA sites and are within suitable distances of adjacent ACA sites; in this sense, MASTER-seq is analogous to reduced representation bisulfite sequencing (RRBS) (Meissner et al., 2005), which provides a quantitative readout for a subset of DNA m5C sites. (2) For absolute (but not relative) quantification, cleavage efficiencies need to be normalized by their counterparts in methylation deficient backgrounds. While mutants are not always available, these can potentially be replaced by synthetic *in vitro*-transcribed RNA, by demethylase treating the mRNA, or by directly modeling effects due to sequence and structure. (3) Quantifications obtained via MAZTER-seq are tightly connected to the distribution of insert lengths in the sequenced libraries, which can differ from one library to another and complicate between-sample analyses. (4) MazF is not entirely exclusive to ACA sites, and minor levels of cutting are observed also at ACA resembling sequences, such as ACG or AAA (Figure 1C); this can bias quantifications at some sites.

MAZTER-seq, offering a quantitative readout for a subset of sites, thus provides a highly complementary readout to m⁶A-seq, offering a qualitative readout for more sites. The decision as to whether implement MAZTER-seq as a standalone approach or in conjunction with m⁶A-seq and miCLIP should take into account these advantages and vulnerabilities. We anticipate that MAZTER-seq will be of high utility for exploring m⁶A distribution, dynamics, functions, and mechanisms of action.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Yeast cells maintenance and Induction of meiosis
 - Cell Culture and Generation of Stable Cell Lines
 - Deletion of FTO in hESCs
 - Overexpression of FTO and ALKBH5

Figure 7. Quantitative Tracking of m⁶A across Genetic Perturbations and Development

- (A) Scatterplot depicting the correlation between high-confidence m⁶A sites (sites in the high and highest confidence bins) between WT and FTO-KO hESCs cells. The x = y curve is plotted in gray and the regression line in blue. The plotted data represent an average of the triplicates in each condition.
- (B) Empirical distribution function plot depicting the difference between WT and FTO-KO cells (x axis) as a function of the confidence group with which the m⁶A sites are associated.
- (C and D) Analyses as in (A) and (B), respectively, comparing WT HEK293T cells to counterparts in which FTO was overexpressed.
- (E and F) Analyses as in (A) and (B), respectively, comparing WT HEK293T cells to counterparts in which ALKBH5 was overexpressed.
- (G and H) Analyses as in (A) and (B), respectively, comparing mESCs to EBs.
- (I and J) Analyses as in (A) and (B), respectively, comparing mESC WT to counterparts of METTL3 KO.

● METHOD DETAILS

- Reagents
- Antibodies
- DAPI staining
- Cellular Fractionation
- Synthesis of in-vitro transcribed RNA
- mRNA preparation
- MazF digestion
- RNA barcoding
- mRNA m⁶A immunoprecipitation
- Library preparation
- De novo reconstruction of MazF sequence biases
- Systematic perturbation of m⁶A consensus sequence
- LC-MS/MS for determination of m⁶A/A
- Site-specific cleavage and radioactive-labeling followed by ligation-assisted extraction and thin-layer chromatography - SCARLET
- MazF-qPCR
- Read Alignment

● QUANTIFICATION AND STATISTICAL ANALYSIS

- Cleavage efficiency
- IP/INPUT differential analysis
- Detection of putative m⁶A sites and confidence group assignment in yeast
- Estimation of m⁶A antibody-based approaches False Detection Rate
- Assessment of agreement between MAZTER-seq and mass-spectrometry
- Methylation models
- Evolutionary analysis
- Calculation of cleavage efficiencies at known sites in human and mouse samples
- De novo detection of putative m⁶A sites in mouse
- Statistical Analysis and plotting

● DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cell.2019.06.013>.

ACKNOWLEDGMENTS

S.S. is supported by the Israel Science Foundation (543165) and the European Research Council (ERC) (no. 714023).

AUTHOR CONTRIBUTIONS

M.A.G.-C., S.E., and S.S. conceived the experiments. S.E. developed and applied MASTER-SEQ. M.A.G.-C. developed MAZTER-MINE and performed

most analyses in the paper. M.S. validated the model through systematic perturbations in yeast. U.T. and W.R. performed SCARLET analyses. R.S. developed MazF-qPCR. A.B. applied MS to quantify bulk m⁶A levels. S.V., R.N., R.W., J.H.H., and L.L. contributed reagents. S.S. wrote the initial draft of the manuscript. M.A.G.-C. and S.E. edited and reviewed the manuscript with input from all other authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 20, 2018

Revised: April 3, 2019

Accepted: June 6, 2019

Published: June 27, 2019

REFERENCES

- Agarwala, S.D., Blitzblau, H.G., Hochwagen, A., and Fink, G.R. (2012). RNA methylation by the MIS complex regulates a cell fate decision in yeast. *PLoS Genet.* 8, e1002732.
- Auguie, B. (2016). gridExtra: Miscellaneous Functions for "Grid" Graphics. <https://rdrr.io/cran/gridExtra/>.
- Bache, S.M., and Wickham, H. (2014). magrittr: A Forward-Pipe Operator for R. <https://magrittr.tidyverse.org/>.
- Ballings, M., and Van den Poel, D. (2013). AUC: Threshold independent performance measures for probabilistic classifiers. <https://rdrr.io/cran/AUC/>.
- Bolstad, B.M. (2013). preprocessCore: A collection of pre-processing functions. <https://github.com/bmbolstad/preprocessCore>.
- Brar, G.A., Yassour, M., Friedman, N., Regev, A., Ingolia, N.T., and Weissman, J.S. (2012). High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* 335, 552–557.
- Brillante, N., Göhringer, M., Lindenhofer, D., Toth, U., Rossmanith, W., and Hartmann, R.K. (2016). Substrate recognition and cleavage-site selection by a single-subunit protein-only RNase P. *Nucleic Acids Res.* 44, 2323–2336.
- Chen, K., Wei, Z., Zhang, Q., Wu, X., Rong, R., Lu, Z., Su, J., de Magalhães, J.P., Rigden, D.J., and Meng, J. (2019). WHISTLE: a high-accuracy map of the human N6-methyladenosine (m6A) epitranscriptome predicted using a machine learning approach. *Nucleic Acids Res.* 47.
- Chen-Kiang, S., Nevins, J.R., and Darnell, J.E., Jr. (1979). N-6-methyl-adenosine in adenovirus type 2 nuclear RNA is conserved in the formation of messenger RNA. *J. Mol. Biol.* 135, 733–752.
- Cheng, Z., Otto, G.M., Powers, E.N., Keskin, A., Mertens, P., Carr, S.A., Jovanovic, M., and Brar, G.A. (2018). Pervasive, Coordinated Protein-Level Changes Driven by Transcript Isoform Switching during Meiosis. *Cell* 172, 910–923.
- Clancy, M.J., Shambaugh, M.E., Timpte, C.S., and Bokar, J.A. (2002). Induction of sporulation in *Saccharomyces cerevisiae* leads to the formation of N6-methyladenosine in mRNA: a potential mechanism for the activity of the IME4 gene. *Nucleic Acids Res.* 30, 4509–4518.
- Darnell, R.B., Ke, S., and Darnell, J.E., Jr. (2018). Pre-mRNA processing includes N⁶ methylation of adenosine residues that are retained in mRNA exons and the fallacy of "RNA epigenetics". *RNA* 24, 262–267.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amarilio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485, 201–206.
- Fu, Y., Dominissini, D., Rechavi, G., and He, C. (2014). Gene expression regulation mediated through reversible m⁶A RNA methylation. *Nat. Rev. Genet.* 15, 293–306.
- Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A., et al. (2013). Derivation of novel human ground state naïve pluripotent stem cells. *Nature* 504, 282–286.
- Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A.A., Kol, N., Salmon-Divon, M., Hershkovitz, V., Peer, E., Mor, N., Manor, Y.S., et al. (2015). Stem cells. m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science* 347, 1002–1006.
- Grozhik, A.V., and Jaffrey, S.R. (2018). Distinguishing RNA modifications from noise in epitranscriptome maps. *Nat. Chem. Biol.* 14, 215–225.
- Harper, J.E., Miceli, S.M., Roberts, R.J., and Manley, J.L. (1990). Sequence specificity of the human mRNA N6-adenosine methylase in vitro. *Nucleic Acids Res.* 18, 5735–5741.
- Herzog, V.A., Reichholz, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T.R., Wlotzka, W., von Haeseler, A., Zuber, J., and Ameres, S.L. (2017). Thiol-linked alkylation of RNA to assess expression dynamics. *Nat. Methods* 14, 1198–1204.
- Horowitz, S., Horowitz, A., Nilsen, T.W., Munns, T.W., and Rottman, F.M. (1984). Mapping of N6-methyladenosine residues in bovine prolactin mRNA. *Proc. Natl. Acad. Sci. USA* 81, 5667–5671.
- Imanishi, M., Tsuji, S., Suda, A., and Futaki, S. (2017). Detection of N⁶-methyladenosine based on the methyl-sensitivity of MazF RNA endonuclease. *Chem. Commun. (Camb.)* 53, 12930–12933.
- Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y.-G., and He, C. (2011). N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* 7, 885–887.
- Ke, S., Alemu, E.A., Mertens, C., Gantman, E.C., Fak, J.J., Mele, A., Haripal, B., Zucker-Scharff, I., Moore, M.J., Park, C.Y., et al. (2015). A majority of m6A residues are in the last exons, allowing the potential for 3' UTR regulation. *Genes Dev.* 29, 2037–2053.
- Knuckles, P., and Bühl, M. (2018). Adenosine methylation as a molecular imprint defining the fate of RNA. *FEBS Lett.* 592, 2845–2859.
- Lengner, C.J., Gimelbrant, A.A., Erwin, J.A., Cheng, A.W., Guenther, M.G., Welstead, G.G., Alagappan, R., Frampton, G.M., Xu, P., Muffat, J., et al. (2010). Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations. *Cell* 141, 872–883.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Linder, B., Grozhik, A.V., Olarerin-George, A.O., Meydan, C., Mason, C.E., and Jaffrey, S.R. (2015). Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods* 12, 767–772.
- Liu, N., and Pan, T. (2015). Probing RNA Modification Status at Single-Nucleotide Resolution in Total RNA. *Methods Enzymol.* 560, 149–159.
- Liu, N., Parisien, M., Dai, Q., Zheng, G., He, C., and Pan, T. (2013). Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA* 19, 1848–1856.
- Lorenz, R., Bernhart, S.H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 6, 26.
- Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A.V., Patil, D.P., Linder, B., Pickering, B.F., Vasseur, J.-J., Chen, Q., et al. (2017). Reversible methylation of m⁶A_m in the 5' cap controls mRNA stability. *Nature* 541, 371–375.
- Mauer, J., Sindelar, M., Despic, V., Guez, T., Hawley, B.R., Vasseur, J.-J., Rentmeister, A., Gross, S.S., Pellizzoni, L., Debart, F., et al. (2019). FTO controls reversible m⁶Am RNA methylation during snRNA biogenesis. *Nat. Chem. Biol.* 15, 340–347.
- Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* 33, 5868–5877.
- Merkurjev, D., Hong, W.-T., Iida, K., Oomoto, I., Goldie, B.J., Yamaguti, H., Ohara, T., Kawaguchi, S.-Y., Hirano, T., Martin, K.C., et al. (2018). Synaptic

- N⁶-methyladenosine (m⁶A) epitranscriptome reveals functional partitioning of localized transcripts. *Nat. Neurosci.* 21, 1004–1014.
- Meyer, K.D., and Jaffrey, S.R. (2014). The dynamic epitranscriptome: N⁶-methyladenosine and gene expression control. *Nat. Rev. Mol. Cell Biol.* 15, 313–326.
- Meyer, K.D., and Jaffrey, S.R. (2017). Rethinking m⁶A Readers, Writers, and Erasers. *Annu. Rev. Cell Dev. Biol.* 33, 319–342.
- Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* 149, 1635–1646.
- Neuwirth, E. (2014). RColorBrewer: ColorBrewer palettes. <https://CRAN.R-project.org/package=RColorBrewer>.
- Nevins, J.R., and Darnell, J.E., Jr. (1978). Steps in the processing of Ad2 mRNA: poly(A)+ nuclear sequences are conserved and poly(A) addition precedes splicing. *Cell* 15, 1477–1493.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- R Development Core Team (2017). R: A language and environment for statistical computing (R Foundation for Statistical Computing).
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.
- Safra, M., Sas-Chen, A., Nir, R., Winkler, R., Nachshon, A., Bar-Yaacov, D., Erlacher, M., Rossmanith, W., Stern-Ginossar, N., and Schwartz, S. (2017). The m(1)A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* 551, 251–255.
- Schwalb, B., and Tresch, A. (2015). LSD: Lots of Superior Depictions. <https://rdrr.io/cran/LSD/>.
- Schwartz, S. (2016). Cracking the epitranscriptome. *RNA* 22, 169–174.
- Schwartz, S., Agarwala, S.D., Mumbach, M.R., Jovanovic, M., Mertins, P., Shishkin, A., Tabach, Y., Mikkelsen, T.S., Satija, R., Ruvkun, G., et al. (2013). High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell* 155, 1409–1421.
- Schwartz, S., Bernstein, D.A., Mumbach, M.R., Jovanovic, M., Herbst, R.H., León-Ricardo, B.X., Engreitz, J.M., Guttman, M., Satija, R., Lander, E.S., et al. (2014a). Transcriptome-wide mapping reveals widespread dynamically-regulated pseudouridylation of ncRNA and mRNA. *Cell* 159, 148–162.
- Schwartz, S., Mumbach, M.R., Jovanovic, M., Wang, T., Maciag, K., Bushkin, G.G., Mertins, P., Ter-Ovanesyan, D., Habib, N., Cacchiarelli, D., et al. (2014b). Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell Rep.* 8, 284–296.
- Wagih, O. (2017). ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* 33, 3645–3647.
- Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., et al. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 505, 117–120.
- Warnes, G.R., Bolker, B., and Lumley, T. (2014). gtools: Various R programming tools. <https://CRAN.R-project.org/package=gtools>.
- Wei, C.M., and Moss, B. (1977). Nucleotide sequences at the N6-methyladenosine sites of HeLa cell messenger ribonucleic acid. *Biochemistry* 16, 1672–1676.
- Wei, J., Liu, F., Lu, Z., Fei, Q., Ai, Y., He, P.C., Shi, H., Cui, X., Su, R., Klungland, A., et al. (2018). Differential m⁶A, m⁶A_m, and m¹A Demethylation Mediated by FTO in the Cell Nucleus and Cytoplasm. *Mol. Cell* 71, 973–985.
- Wickham, H. (2010). *ggplot2: Elegant Graphics for Data Analysis* (Springer).
- Wickham, H., and Henry, L. (2016). tidy: Easily Tidy Data with 'spread (')' and 'gather (')'Functions. <https://tidyverse.org/>.
- Wickham, H., François, R., Henry, L., and Müller, K. (2015). dplyr: A grammar of data manipulation. <https://dplyr.tidyverse.org/>.
- Yue, Y., Liu, J., and He, C. (2015). RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation. *Genes Dev.* 29, 1343–1355.
- Zeng, Y., Wang, S., Gao, S., Soares, F., Ahmed, M., Guo, H., Wang, M., Hua, J.T., Guan, J., Moran, M.F., et al. (2018). Refined RIP-seq protocol for epitranscriptome analysis with low input materials. *PLoS Biol.* 16, e2006092.
- Zhang, Y., Zhang, J., Hoeftlich, K.P., Ikura, M., Qing, G., and Inouye, M. (2003). MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in *Escherichia coli*. *Mol. Cell* 12, 913–923.
- Zhao, B.S., Nachtergaelie, S., Roundtree, I.A., and He, C. (2018). Our views of dynamic N⁶-methyladenosine RNA methylation. *RNA* 24, 268–272.
- Zheng, G., Dahl, J.A., Niu, Y., Fedorcsak, P., Huang, C.-M., Li, C.J., Vågbo, C.B., Shi, Y., Wang, W.-L., Song, S.-H., et al. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* 49, 18–29.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti-m ⁶ A antibody	Synaptic Systems	Cat# 202003; RRID: AB_2279214
Mouse polyclonal Anti-ALKBH5	Abcam	Cat# ab69325; RRID: AB_1267666
Rabbit monoclonal anti-FTO	Abcam	Cat# ab124892; RRID: AB_10972698
Chemicals, Peptides, and Recombinant Proteins		
mRNA interferase-MazF enzyme	Takara Bio	Cat# 2415A
DAPI staining	ThermoScientific	Cat# D3571
Critical Commercial Assays		
PARIS kit	Invitrogen	Cat# AM1921
Deposited Data		
Raw and analyzed data	This paper	GEO: GSE122961
m ⁶ A-seq yeast data	Schwartz et al., 2013	GEO: GSE51583
Experimental Models: Cell Lines		
Human Fetus HEK293T cells (female)	American Type Culture Collection (ATCC)	ATCC CRL-11268
Murine Pluripotent Embryonic Stem Cells (ESCs) WT (male)	Geula et al., 2015	N/A
Murine Pluripotent Embryonic Stem Cells (ESCs) KO Mettl3 ^{-/-} (male)	Geula et al., 2015	N/A
Experimental Models: Organisms/Strains		
S. cerevisiae WT (SAY821) MAT a/α lys2/lys2 ho::LYS2/ho::LYS2	Agarwala et al., 2012	N/A
S. cerevisiae ndt80Δ/Δ (SAY841) MAT a/α lys2/ lys2 ho::LYS2/ho::LYS2 ndt80::LEU2/ndt80::LEU2	Agarwala et al., 2012	N/A
S. cerevisiae lme4Δ/Δ (SAY966) MAT a/α lys2/lys2 ho::LYS2/ho::LYS2 ndt80::LEU2/ndt80::LEU2 lme4:: HIS3/lme4::HIS3	Schwartz et al., 2013	N/A
Oligonucleotides		
Synthetic RNA oligos for MAZTER-seq calibration (IVT1-2), see Table S7	This paper	N/A
Oligonucleotides for library construction, see Table S7	This paper	N/A
MazF-qPCR sequences, see Table S7	This paper	N/A
Oligonucleotides for SCARLET analysis, see Table S7	This paper	N/A
Oligos for MazF biases check analysis, see Table S7	This paper	N/A
Oligos for Systematic perturbation of m ⁶ A consensus sequence analysis, see Table S7	This paper	N/A
Recombinant DNA		
pcDNA3.1+ /C-(K)DYK plasmids encoding full-length FTO (NM_001080432)	GenScript	Cat#: OHu24899D
pcDNA3.1+ /C-(K)DYK plasmids encoding full-length ALKBH5	GenScript	Cat#: OHu31092D
pEGFP-N3	SignaGen	Cat#: SL100688
CRISPR/Cas9 plasmid	Addgene	Cat#: pV1382
Software and Algorithms		
STAR	Dobin et al., 2013	https://github.com/alexdobin/STAR
SAMtools	Li et al., 2009	http://samtools.sourceforge.net/

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
BEDtools	Quinlan and Hall, 2010	https://bedtools.readthedocs.io/en/latest/
R	R Development Core Team, 2017	https://www.r-project.org/
MAZTER-MINE	This paper	https://github.com/SchwartzLab/mazter_mine
ViennaRNA	Lorenz et al., 2011	https://www.tbi.univie.ac.at/RNA/

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Schraga Schwartz (schwartz@weizmann.ac.il).

EXPERIMENTAL MODEL AND SUBJECT DETAILS**Yeast cells maintenance and Induction of meiosis**

S. cerevisiae cells (WT, ndt80 Δ/Δ , lme4 Δ/Δ) were grown in YPD (2% dextrose) at 30°C. To induce synchronous meiotic entry, cells were grown for 24 h in 1% yeast extract, 2% peptone, 4% dextrose at 30°C, diluted in BYTA (1% yeast extract, 2% tryptone, 1% potassium acetate, 50 mM potassium phthalate) to OD600 = 0.2 and grown for another 16 h at 30°C, 200 rpm. Cells were then washed once with water and re-suspended in SPO (0.3% potassium acetate) at OD600 = 2.0 and incubated at 30°C at 190 rpm. Cells were isolated from SPO at the indicated times and collected by 2 min centrifugation at 3000 g. Pellets were snap frozen and stored at -80 for RNA extraction.

Cell Culture and Generation of Stable Cell Lines

Maintenance of naive murine pluripotent cells (male) was conducted in FBS free N2B27-based media (Geula et al., 2015). 500mL of N2B27 media was generated by including: 240 mL DMEM/F12 (Biological Industries – HEPES free, custom made), 240 mL Neurobasal (Invitrogen; 21103), 5 mL N2 supplement (Invitrogen; 17502048 or in house prepared), 5 mL B27 supplement (Invitrogen; 17504044), 1 mM glutamine (Invitrogen), 1% nonessential amino acids (Invitrogen), 0.1 mM β -mercaptoethanol (Sigma), penicillin-streptomycin (Invitrogen), 5 mg/mL BSA (Sigma). Naive conditions for murine ESC included 10 μ g recombinant human LIF (Peprotech) and small-molecule inhibitors CHIR99021 (CH, 1-3 μ M- Axon Medchem) and PD0325901 (PD, 1 μ M - TOCRIS) termed 2i. Primed N2B27 media for murine cells (EpiSCs) contained 8 ng/mL recombinant human bFGF (FGF2) (Peprotech), 20 ng/mL recombinant human Activin (Peprotech), and 1% Knockout serum replacement (KSR- Invitrogen). For FBS/LIF growth conditions, mESCs expansion was carried out in 500 mL of High-glucose DMEM (ThermoScientific), 15% USDA certified fetal bovine serum (FBS- Biological Industries), 1 mM L-Glutamine (Biological Industries), 1% nonessential amino acids (Biological Industries), 0.1 mM β -mercaptoethanol (Sigma), 1% penicillin-streptomycin (Biological Industries), 1% Sodium-Pyruvate (Biological Industries), 10 μ g recombinant human LIF (Peprotech). Cells were maintained in 20% O₂ conditions on irradiation inactivated mouse embryonic fibroblast (MEF) feeder cells, and were passaged following 0.25% trypsinization. For RNA extraction, cells were grown on Gelatin for three passages in FBS free N2B27-based media (Gafni et al., 2013). Briefly, 500mL of N2B27 media was produced by including: 250 mL DMEM:F12 (ThermoScientific), 250 mL Neurobasal (ThermoScientific), 5 mL N2 supplement (Invitrogen; 17502048 or in-house prepared), 5 mL B27 supplement (Invitrogen; 1750404), 1 mM L-Glutamine (Biological Industries), 1% nonessential amino acids (Biological Industries), 0.1 mM β -mercaptoethanol (Sigma), penicillin-streptomycin (Biological Industries). Naive conditions for murine ESCs included 10 μ g recombinant human LIF (Peprotech) and small-molecule inhibitors CHIR99021 (CH, 3 μ M- Axon Medchem) and PD0325901 (PD, 1 μ M - Axon Medchem) termed 2i. For *in vitro* embryoid bodies (EBs) formation, roughly 5 million mESCs were disaggregated with trypsin and transferred to non-adherent suspension culture dishes, and cultured in MEF medium (DMEM supplemented with 1% L-Glutamine, 1% Non-essential amino acids, 1% penicillin-streptomycin, 1% Sodium-Pyruvate and 15% FBS – does not contain LIF or 2i) for 9 days. Media replacement was carried out every 2 days.

Deletion of FTO in hESCs

Human WT female WIBR3 hESC line (Lengner et al., 2010) was used to generate FTO KO clones. Exon1 of FTO locus was targeted with CRISPR/Cas9 with the following sgRNAs: sense: CACCG CTCTCGTTCTCGGCAGTCG; anti-sense: AAAC CGACTGCCGAG GAACGAGAGC. The following primers were used to detect potential mutant clones with HRM: Forward: GCAGGACGCTGAGA GAACTACATGC; Reverse: CGCCCATCAAACCAAGCCCTTC (expected WT size 218bp). Clones were then validated with western blot analysis to confirm loss of FTO protein detection and also sent for sequencing to characterize the nature of the mutations induced on both FTO alleles. Primers used to obtain the latter were: Forward: CAGCTGTCGGACCTGGGAAATT; Reverse: CATCCCCTC TAATATCTCCAGGTCCCT (expected WT size 412bp). Clone I1 was confirmed to have 22 bp deletion in exon 1 of both alleles which yields a truncated protein derivative of only 9 amino acids (out of 505 total in WT protein), and was used for further analysis alongside

its isogenic WIBR3 WT donor hESC line. The following conditions were used to expand and genetically target hESCs: DMEM-F12 (Invitrogen 10829) supplemented with 15% Knockout Serum Replacement (Invitrogen 10828-028), 1 mM GlutaMAX (Invitrogen), 1% nonessential amino acids (Invitrogen), 0.1 mM β -mercaptoethanol (Sigma), and 8 ng/mL bFGF (Peprotech) and 1 ng/mL recombinant human TGF β 1 (Peprotech). Cultures were passaged every 5–7 days either manually or by trypsinization (24 h pre and 24 h after addition of ROCK inhibitor at 5–10 μ M concentration). For transfection of hiPSC and hESC lines, cells were cultured in Rho kinase (ROCK) inhibitor (Y27632) 24 h before electroporation. Primed human ESC and iPSC cells were harvested with 0.25% trypsin-EDTA solution (Invitrogen), and cells resuspended in PBS were transfected with DNA constructs (Gene Pulser Xcell System; Bio-Rad; 250 V, 500 μ F, 0.4 cm cuvettes). Cells were subsequently plated on MEF feeder layers (DR4 MEFs) in hESC medium supplemented with ROCK inhibitor for the first 24 h.

Overexpression of FTO and ALKBH5

Human (female) HEK293T cells (American Type Culture Collection (ATCC); passage numbers 8–15 were plated in 10 cm plates at 90% confluence. plasmids encoding full-length FTO (NM_001080432, Catalog No: OHu24899D), ALKBH5 (NM_017758, Catalog No: OHu31092D) under a cytomegalovirus promoter were obtained from GenScript. As control, cells were transfected with GFP vector (pEGFP-N3) which also enables monitoring the transfection efficiency (86%). The plasmids were transfected into HEK293T cells using PolyJet transfection reagent (SignaGene) with one boost of the plasmid at 24 h. Cells were harvested 48 h after transfection.

METHOD DETAILS

Reagents

mRNA interferase-MazF enzyme was purchased from Takara Bio (Cat#2415A).

Antibodies

Rabbit anti-m⁶A antibody for mRNA immunoprecipitation was purchased from Synaptic Systems (Cat# 202003). Anti-FTO and anti-AlkBH5 were purchased from Abcam (ab124893 and ab69325 respectively).

DAPI staining

To observe the progression in cells meiosis, a sample of cells were taken at each time point during sporulation (0 – 8 h) for DAPI staining. Cells were first fixed by Formaldehyde (J.T Baker, UN2209) followed by incubation with DAPI (ThermoScientific, D3571) reagent at 4°C for 2 h. Staining was observed by Olympus IX73 Fluorescent microscope system.

Cellular Fractionation

Human (female) HEK293T cells (American Type Culture Collection (ATCC)); passage numbers 8–15 were plated in 10 cm plates at 100% confluence. Fractionation was performed using PARIS kit (Invitrogen, AM1921) according to manufacturer's protocol.

Synthesis of in-vitro transcribed RNA

Two synthetic RNA fragments (IVT1 and IVT2, [Table S7](#)), each comprising a 103 nt long sequence with a single ACA in the center were *in vitro* transcribed from dsDNA templates, either in the presence of ATP or N6-methyl-ATP, using MaxiScriptT7 kit (Invitrogen, AM1320). Purified m⁶A-containing products were serially diluted in non-m⁶A containing products.

mRNA preparation

Yeast total RNA samples were prepared by MasterPure Yeast RNA extraction kit (Lucigen, MPY03100). For mouse ESCs cells, total RNA was extracted using Nucleozol (Macherey-Nagel, 740404.200). Enrichment of polyadenylated RNA from total RNA was performed using Oligo(dT) dynabeads mRNA-DIRECT kit (Thermo Scientific, 61012) for small mRNA amounts. Large mRNA amounts (> 1 μ g) were prepared by GenElute mRNA miniprep kit (Sigma, MRN70). All kits procedures were conducted according to the manufacturer's protocol.

MazF digestion

100 ng of poly A selected RNA was first heat denatured at 70°C for 2 min and placed on ice. Each sample was supplemented with 4 μ L 5x buffer, 0.8 μ L DMSO, 0.5 μ L RNase inhibitor (NEB, M0314L) and 20 units of MazF enzyme (TakaRa, 2415A). Reactions were incubated at 37°C for 30 min and stopped by placing on ice and RNA cleanup. RNA was purified by adding 3 \times volume Buffer RLT (QIAGEN, 79216) and 1 \times volume ethanol, precipitating onto MyOne SILANE Dynabeads (Invitrogen, 37002D) beads, washing twice in 80% ethanol, and eluting in water.

RNA barcoding

MazF cleaved RNA was first end-repaired using 1:1 ratio of Fast AP buffer (Thermo Scientific EF0652) and T4 PNK buffer (NEB, M0201L), in the presence of Turbo DNase 1ul, Fast AP 3ul and T4 PNK 4ul. Reactions were cleaned of enzymes by adding 3 \times volume Buffer RLT (QIAGEN, 79216) and 1 \times volume ethanol, precipitating onto MyOne SILANE Dynabeads (Invitrogen, 37002D) beads,

washing twice in 80% ethanol, and eluting in water. We combined dephosphorylated RNA with 20 pmol RNA adaptor 3iLL (Table S7), denatured at 70°C for 2 min, then snap-cooled by transferring to ice. Ligation reaction proceeded by addition of T4 RNA Ligase 1 (New England Biolabs, M0437M) at 23°C for 75 min. Ligated RNA was purified by adding 3 × volume Buffer RLT (QIAGEN) and 0.85 × volume ethanol, precipitating onto SILANE beads, washing twice in 80% ethanol, and eluting in water. Ligated RNA samples were pooled together. In applications not involving m⁶A-IP we proceeded directly to 'library preparation' section. In ones involving IP, the pool was divided into Input sample (5% of material) and IP (95%).

mRNA m⁶A immunoprecipitation

m⁶A-IPs were performed essentially as described in ref. (Safra et al., 2017; Schwartz et al., 2013), using a total of 1-2 µg of poly(A) selected MazF-treated starting material, following pooling. 25 µL of protein-G magnetic beads were washed and resuspended in 200 µL of 1x IPP buffer (NaCl 300mM, NP-40 0.2%, Tris-HCl 20mM), and tumbled with 3 µL of affinity purified anti-m⁶A polyclonal antibody (Synaptic Systems, 202003) at room temperature for 30 min. Following 2 washes in IPP buffer, IP sample RNA was added to the antibody-bead mixture, and incubated for 2 h at 4°C. The RNA was then washed twice in 200 µL of IPP buffer, twice in low-salt IPP buffer (50 mM NaCl, 0.1% NP-40, 10mM Tris-HCl, pH 7.5), and twice in high-salt IPP buffer (500 mM NaCl, 0.1% NP-40, 10mM Tris-HCl, pH 7.5), and eluted in 30 µL RLT (QIAGEN). To purify the RNA, 20 µL MyOne Silane Dynabeads (Life Technologies) were washed in 100 µL RLT, resuspended in 30 µL RLT, and added to the eluted RNA. 60 µL 100% ethanol was added to the mixture, the mixture attached to the magnet and the supernatant discarded. Following two washes in 100 µL of 70% ethanol, the RNA was eluted from the beads in 160 µL IPP buffer. Eluted RNA was subjected to an additional round of IP, by reincubating it with protein-A magnetic beads coupled to anti-m⁶A antibody, followed by washes, elution from the protein-A beads and purification as above, followed by elution from the MyOne silane dynabeads in 10 µL H2O.

Library preparation

A DNA primer (rTdT cDNA primer, Table S7), complementary to the RNA ligated adaptor (3iLL, Table S7), was used for cDNA synthesis by SuperScript III Reverse transcriptase (Thermo Fisher, 18080093). The reaction was performed according to manufacturer instructions for 1 h at 50°C, without heat-inactivation to preserve RNA-cDNA hybrids. The remaining primers left in the reaction were digested by addition of 3 µL ExoSAP-IT (Affymetrix, 75001) and incubation at 37°C for 12 min. ExoSAP activity was stopped by addition of EDTA. RNA was degraded by adding 2.5 µL of NaOH 1M and incubation at 70°C for 12 min. The base was neutralized with the addition of 1M HCl, then cDNA was cleaned using MyOne silane beads as described in the previous section. A second adaptor was added to the cDNA by ligating 50 pmol 5iLL-22 DNA adaptor (Table S7) with T4 RNA Ligase 1 (NEB, M0437M), DMSO 4%, ATP 1mM and PEG8000 21% at 23°C for 3 h. Following clean up with MyOne silane beads, we amplified the cDNA library for 13 cycles using Library PCR primers F and R (Table S7) and KAPA HiFi HotStart ReadyMix (KAPA Biosystems KK2601). Amplified libraries were cleaned with AMPure XP beads (Agencourt, A63881), quantified using Qubit (Life Technologies) and the distribution of library size was determined using TapeStation (Agilent Technologies). Libraries were sequenced on an Illumina NextSeq 500 machine.

De novo reconstruction of MazF sequence biases

RNA with the following sequence CAAAAGGAAGGTGGAAAAGANNNNACANNNAAGCAAAATGACTTGCTAGATCTTGGC, and a control sequence lacking an ACA site CAAAAGGAAGGTGGAAAAGATCGTTGTGTAAAGCAAAATGACTTGCTAGATCTTGGC, were *in vitro* synthesized using MEGAscript kit (Thermo). DNase treatment (using Turbo DNase, Thermo) was carried out to eliminate DNA traces. Purified RNA was either MazF treated or left untreated. RNA was reverse transcribed using superscript III (Thermo) kit with specific primer (CACGACGCTCTCCGATCTCGGCCAAGATCTAGCAAGTCA). Nested cDNA amplification was then performed in two rounds, beginning with 10 cycles using AGACGTGTGCTCTCCGATCTCCAAAAGGAAGGTGGAAAAG and CACGACGCTCTCCGATCTCGGCCAAGATCTAGCAAGTCA followed by 20 rounds of amplification using standard illumina primers.

Systematic perturbation of m⁶A consensus sequence

To systematically modify sequences surrounding an m⁶A site in the 3' UTR of the GIP2 gene (chr5:268382 in the sk1 genome), yeast were co-transformed with a CRISPR/Cas9 plasmid (pV1382, Addgene) carrying the following gRNA sequence: AAAAG GAAGGTGATGAAGAA, and with a set of 256 repair templates using the following DNA primer: GGCAGCATGAAGGCAACCAAAG GAAGGTGGAAAAGANRRGACANNAAAAGCAAAATGACTTGCTAGATCTTGGCCTGAG. Yeast were grown and collected during meiosis prophase, RNA was extracted using MasterPure Yeast RNA Purification kit (Lucigen). Harsh DNase treatment was applied using Turbo DNase (Thermo) to avoid genomic DNA contamination. Specific amplification was done after reverse transcription with random hexamer and superscript III (Thermo) using the following primers: AGACGTGTGCTCTCCGATCTCCAAAAGGAAGGTG GAAAAG and CACGACGCTCTCCGATCTCGGCCAAGATCTAGCAAGTCA, followed by an amplification using standard illumina adaptors. Relative abundances of each variant were quantified. We required at least 20 reads per variant in the Input (non-MazF treated) libraries; In addition, to avoid potential confoundment with MazF cleavage preferences, we further filtered out all variants harboring a 'G' at position +3. To each sequence variant we applied the linear model. Given that each of the sequences is embedded precisely in the same relative position within the gene, and in a fixed genomic environment, we used a uniform estimate of these in the models across all variants.

LC-MS/MS for determination of m⁶A/A

To analyze nucleotide composition, 400 ng of double selected polyA RNA fractions were digested with 2 units of P1 nuclease (US biological) at 50°C in 50 mM ammonium acetate buffer pH 5.3, with 5mM zinc chloride for 2 h. Nucleotides were dephosphorylated by addition of 5 units of CIP (New England Biolabs) for another 2 h at 37°C, and then diluted 1:10 in acetonitrile. The samples were dried by acetonitrile evaporation in speedvac. The residue of each sample was re-dissolved in 198 µL of 0.01% formic acid. Two µL of 1 µg/mL 7-deaza-A were added as internal standard. The mixtures were intensively vortexed (0.5min), centrifuged (21,000rpm; 5min), and passed through 0.22-µm PVDA filters (Millex GV) to 250 µL inserts of LC-MS vials. The LC-MS/MS instrument consisted of an Acquity I-class UPLC system (Waters) and Xevo TQ-S triple quadrupole mass spectrometer (Waters) equipped with an electrospray ion source and operated in positive ion mode was used for analysis of nucleosides. MassLynx and TargetLynx software (version 4.1, Waters) were applied for the acquisition and analysis of data. Chromatographic separation was done on a 100mm × 2.1mm internal diameter, 1.8-µm UPLC HSS T3 column equipped with 50mm × 2.1mm internal diameter, 1.8-µm UPLC HSS T3 pre-column (both Waters Acquity) with mobile phases A (0.01% formic acid) and B (50% aqueous acetonitrile with 0.01% formic acid) at a flow rate of 0.2mL/min and column temperature 25°C. A gradient was used as follows: the column was held at 0%B for 1 min, then a non-linear increase (curve 8) to 35%B from 1 to 18min, then a non-linear increase (curve 8) to 100%B 18-18.2 min, held at 100%B 18.2-19min, back to 0% B 19-20 min and equilibration at 0% B for additional 5min. Samples kept at 7°C were automatically injected in a volume of 1 or 3 µL, to get non-saturated A and m⁶A signals, respectively. Retention times were 9.7, 11.4, 11.8, and 13.8 min for 7-deaza-A, G, A, and N6-Me-A respectively. For mass spectrometry, argon was used as the collision gas with a flow of 0.10mL/min. The capillary voltage was set to 2.67kV, source temperature 150°C, desolvation temperature 400°C, cone gas flow 150L/h, desolvation gas flow 800L/h.

Nucleoside concentration was calculated using a standard curve of the relevant nucleotide concentration in each sample. Standard curves included increasing concentration of all measured nucleosides ranging from 0-1000ng/mL that were positioned at the beginning and at the end of each run. All the calculated values for the different nucleosides in each sample fell within the standard curve range. The compounds were detected in positive mode as multiple-reaction monitoring, with the following parameters: 267.1 > 118.1 and 267.1 > 135.0 m/z (collision energy CE 57 and 16 eV respectively) for 7-deaza-A, 284.2 > 152.1 m/z (CE 14eV) for G, 268.1 > 136.1 m/z (CE 15 eV) for A, and 282.1 > 123.1 and 282.1 > 150.1 m/z (CE 40 and 25 eV respectively) for N6-Me-A.

Site-specific cleavage and radioactive-labeling followed by ligation-assisted extraction and thin-layer chromatography - SCARLET

Site-specific cleavage and radioactive-labeling followed by ligation-assisted extraction and thin-layer chromatography (TLC) (SCARLET) was carried out as previously described (Liu and Pan, 2015; Liu et al., 2013). In detail, 1-1.5 µg mRNA with 3 pmol chimeric cleavage oligonucleotide (Table S7) in 3 µl 30 mM Tris·Cl, pH 7.5, were denatured for 1 min at 95°C, followed by 3 min annealing at room temperature. The annealing reaction was set on ice and 2 µl of a mixture containing 1 × T4-polynucleotide kinase buffer A (Thermo Scientific), 0.5 U/µl RNase H (enzymatics), 0.5 U/µl FastAP thermosensitive alkaline phosphatase (Thermo Scientific), were added. The reaction was incubated for 1 h at 44°C and subsequently terminated at 75°C for 5 min. For labeling, 1 µl fresh [γ -32P]ATP (20 mCi/mL, 5000 Ci/mmol; Hartmann Analytics) and 1 µl T4 polynucleotide kinase (10 U/µl; Thermo Scientific) were added and the reactions incubated first at 37°C and then at 50°C, for 30 min each; labeling reactions were again terminated at 75°C for 5 min. 4 pmol ligation (splint) oligonucleotide (Table S7) and 5 pmol of the previously described 116-mer DNA oligonucleotide (Liu et al., 2013) premixed in 1 µl were added, incubated 3 min at 75°C and annealed for 3 min at room temperature. 2 µl 5 × ligation mixture (2 × T4-polynucleotide kinase buffer A, 68% DMSO, 0.4 mM ATP, 2.4 U/µl T4 DNA ligase (Thermo Scientific)) were added and the reaction incubated for 4 h at 37°C. Thereafter the reaction was stopped by transferring the whole volume to a dried pellet of urea and EDTA prepared by drying 7 µl stop/loading buffer (Brillante et al., 2016) and 0.5 µl 0.5 M EDTA in a vacuum concentrator. 1 µl RNase mixture (100 U/µl RNase T1 (Roche), 0.1 mg/mL RNase A (Sigma)) was added and the RNA digested over-night at 37°C. Next day the samples were resolved by electrophoresis in a 10% denaturing (7 M urea) polyacrylamide gel, visualized by phosphorimaging, the band corresponding to the 116-mer DNA oligonucleotide ligated to the labeled nucleoside of interest cut from the gel and eluted by soaking in 0.5 M ammonium acetate, 0.1% SDS, 1 mM EDTA overnight. Next day the supernatant was extracted with phenol:chloroform, the sample precipitated by ethanol using linear polyacrylamide as carrier, dissolved in 4 µl 60 mM ammonium acetate pH 5.3, 1 µl nuclease P1 (0.5 U/µl) added, and the sample subjected to complete hydrolysis at 37°C for 1 h. 1-2 µl hydrolysate were spotted on cellulose-coated TLC glass plates (CEL 300-10; Macherey-Nagel) and the TLC developed in 2-propanol:HCl (32%):H₂O (70:15:15). Resolved 5'-monophosphate nucleosides were detected by phosphorimaging and quantified using ImageQuant TL 8 image analysis software (GE Healthcare).

To prioritize sites for validation via SCARLET, we combined diverse criteria during different stages of the development of MAZTER-seq, including (1) expression level (SCARLET requires very high amounts of starting material, and we hence prioritized highly expressed genes (ideally with TPM > 100), (2) MAZTER-seq based confidence bins into which the sites were assigned, (3) Predicted methylation score as inferred by the linear model, (4) visual inspection of the locus. In addition to the total of 23 sites at which we quantified m⁶A levels using SCARLET (Figures 2H, 3K, and S3D), we attempted to quantify at 4 additional sites, but failed to obtain informative measurements. In all of these cases, TLC yielded very strong signals - inconsistent with the expression level of the genes harboring the putative interrogated m⁶A sites - that "smeared" substantially across the 'A' and 'm⁶A' areas. We interpreted these

results as reflecting some form of non-specific labeling, and these data points were hence not taken into account in the analyses. For completeness, the full set of sites and their respective oligonucleotides used for their analysis are provided in [Table S7](#).

MazF-qPCR

For the qPCR based readouts of methylation, we designed two primer pairs to interrogate each of two sites. In each case, one ‘test’ primer pair was designed to flank a putative methylation site, and hence upon MazF cleavage will only lead to a product if the site is indeed methylated. The second ‘control’ primer pair was designed to flank an adjacent region in the same gene that did not harbor an ACA site. Readouts were obtained for both pairs either upon MazF digestion or in its absence. The ratio of the abundance of the test versus control primer, in the presence of MazF, was normalized by the corresponding ratio in the absence of MazF treatment. Primer sequences are provided in [Table S7](#).

Read Alignment

Reads were aligned using STAR (V. 2.5.3a) ([Dobin et al., 2013](#)), for the reference genome generation step we used the SK1 reference genome used by ([Schwartz et al., 2013](#)), along with the IVT sequences ([Table S7](#)) as additional chromosomes. Additional parameters used in the alignment step were ‘–alignIntronMax 300–alignMatesGapMax 1000’, all other parameters were set by default. Normalization of paired libraries insert size was performed using in-house python scripts, between pairs of INPUT-IP samples. Alignment sorting and indexing were performed separately with samtools (V. 1.3.1) ([Li et al., 2009](#)). Single-base coverage was retrieved using in-house python scripts and bedtools (V. 2.26.0) ([Quinlan and Hall, 2010](#)), calculating coverage from 3' and 5' positions, and fragment coverage (i.e., from the start of read-1 to the end of read-2).

QUANTIFICATION AND STATISTICAL ANALYSIS

Cleavage efficiency

Strand specific 5' starting, 3' ending, and whole fragment-coverage were retrieved for all ACA motif positions using bedtools and in-house processing scripts. Preliminary cleavage efficiencies were calculated as a ratio of the 3' read-ends or 5' read-ends coverage divided by their respective fragment coverage, sites with fragment coverage less than 15 reads were omitted from the calculation. To identify the minimal distance between an interrogated ACA site and an adjacent ACA site, we employed the following scheme: (1) For each sample retrieve all 5' and 3' cleavage-efficiencies, (2) Iterating from 0 to 200 bp, compute the pairwise correlation between 5' and 3' cleavage efficiencies for sites which adjacent upstream and downstream ACA sites are more distant than the iterate (3) Using the vector of correlation coefficients in step 1, in a step-forward procedure, select the ACA distance at which there is no increase in correlation for the next two increments in ACA distance (4) similarly but separate calculations are done for downstream and upstream directions as defined above, but locking the complementary direction to the preliminary ACA distance calculated in step 3, in this way the threshold of ACA distance is moving in one direction at a time resulting in a ‘minimal upstream distance’ and a ‘minimal downstream distance’ ([Figures S2A and S2B](#)). (5) All ACA sites with share closer proximity to other ACA sites than the minimal downstream and minimal upstream distance, respectively, were subsequently defined as non-available (N/A) measurements. (6) The final cleavage efficiency was defined as the average of the quantifiable 5' and/or 3' cleavage efficiency scores for each site. Note that ‘minimal distances’ is not a fixed value, but a function of the insert size of a library. Hence, this value is computationally inferred, by using a value that optimizes the agreement between the 5' and 3' cleavage efficiencies. For the cases of first and last ACA sites in a gene, the missing closest ACA site distances were replaced with the gene start and end coordinates respectively. Of note, as the SK1 transcriptome annotations do not define UTRs we extended each CDS gene coordinates by 150 bp in the 5' direction and 250 bp in the 3' direction. Due to known MazF motif sequence cleavage bias, and the observation that sites harboring a ‘G’ at position +3 are cleaved with reduced efficiency in comparison to sites lacking it ([Figures S4A–S4C](#)), sites with ‘G’ at position +3 were conservatively omitted from analyses seeking to infer and validate the m⁶A code ([Figures 3A–3B, 3D, 4B, 4C, and 6B–6D](#)).

IP/INPUT differential analysis

IP-m⁶A enrichment calculations were performed using the ‘edgeR’ package ([Robinson et al., 2010](#)). The median of a 50-nt window centered in each ACA sites was used to represent the coverage level of the site, using the previously calculated fragment coverage. Fold-changes were calculated between IP and INPUT treatments in WT strain (m⁶A-Score), and between WT and *ime4Δ/Δ* strains in IP treatment.

Detection of putative m⁶A sites and confidence group assignment in yeast

Putative m⁶A sites were identified using a three-way comparison approach, the comparisons being: (1) sites with less cleavage efficiency in WT than *ime4Δ/Δ* for INPUT samples, (2) sites with less cleavage efficiency in WT than *ime4Δ/Δ* in m⁶A-IP samples, and (3) sites with less cleavage efficiency in m⁶A-IP than INPUT in WT samples. Due to initial differences between libraries in cleavage efficiency distributions (likely originating from differences in the insert sizes of the libraries, see [Discussion](#)), a cross-sample quantile normalization was performed using the ‘preprocessCore’ package ([Bolstad, 2013](#)). Mean difference, log2 Fold-Change (log2FCH) and t test p values were calculated for each site in each of the three comparisons, when fragment coverage at the interrogated position was greater or equal to 15 in all the samples in each comparison.

For confidence group assignment, each site was scored: -2, -1, 0, +1, or +2; depending on whether it was associated with a low or high combination of thresholds for effect size and P value. Specifically, for comparisons 1 and 2, the thresholds for associating a site with a +2 score were mean difference of 0.7 and $p < 0.01$; For a +1 score, the corresponding numbers were 0.5 and 0.05. For comparison 3 thresholds, a +2 score required a log₂FCH of 4 and $p < 0.01$, and a +1 required a log₂FCH of 2 and $p < 0.05$. Scoring significantly for the symmetric but opposite direction would grant a -1 or -2 score respectively (Figure S3A). The scores obtained from the three comparisons were summed for each site resulting in the confidence group assigned to the site. As the number of sites that were assigned to confidence groups 5 and 6 was very small, they were collapsed into confidence group 4. All confidence groups 0 and below were similarly collapsed into the '0' confidence group.

Estimation of m⁶A antibody-based approaches False Detection Rate

To estimate the m⁶A-seq false detection rate, we first used published data to obtain the distances between the m⁶A-seq 'peak' and the nearest consensus sequence (Schwartz et al., 2013). The distributions of distances were inspected as a function of MAZTER-seq derived confidence groups for each of 199 previously identified m⁶A-seq sites that are also quantifiable and had sufficient coverage via MAZTER-seq (Figure S3C). We used the maximum distance between an m⁶A peak and the nearest consensus sequence in sites from confidence group 2 (defined by MAZTER-seq) as a threshold for true positive calls: this threshold was thus conservatively set at 41 nt. Sites with higher distances were highly enriched in lower confidence bins, and this threshold was thus used to estimate the false detection rate of m⁶A-seq.

Assessment of agreement between MAZTER-seq and mass-spectrometry

In this study we performed LC-MS in two replicates for two sporulation experiments. In one we obtained m⁶A:A ratios of ~0.05%, in the other ~0.12% (the differences likely reflect experiment-to-experiment variability in the efficiency of sporulation). To assess whether these numbers were consistent with our estimates on the number and stoichiometry of methylation sites, we performed the following steps: We first applied the linear model predicting methylation levels in a transcriptome wide manner, and provided each site with a 'relative' methylation score. To be able to convert this relative score to an absolute methylation stoichiometry, we made use of the linear model that was obtained based on the sites that had been quantified via SCARLET (Figure 3C). This analysis yielded a total of 6,816 sites with predicted methylation stoichiometries exceeding 0% (Table S4). Given that the contribution of each such methylated position to LC-MS based measurements is not only a function of the stoichiometry but also of its expression levels, we scaled each predicted m⁶A site as well as each A site in each gene based on the relative expression levels in meiotic prophase of the genes harboring them (based on expression data from (Schwartz et al., 2013)). Overall, this analysis predicted an m⁶A: A ratio of 0.026%. Given that only roughly half of the methylated sites in yeast occur at ACA sites - and assuming that the distribution of the stoichiometries of methylation are identical between ACA and non-ACA site - the final m⁶A: A ratio would be predicted to be ~0.05% which overlaps well with one of our two measurements, and is within 2-fold from the other measurement. In contrast, when we performed the above calculation, but on the basis of the ~1,300 originally detected sites (of which 593 are ACA sites), the predicted m⁶A: A ratio was calculated as 0.004.

Methylation models

The variables used for m⁶A modeling were: (1) binarized nucleotide identity for positions -4 to +6 with respect to the interrogated nucleotide, omitting positions 0 to 2 due to the ACA consensus sequence; (2) local minimum free energy within a 60-nt window centered on each ACA site, estimated via the RNAfold function using parameters-noconv-noPS-T 30 in the ViennaRNA package (Lorenz et al., 2011); (3) relative position within the gene (in a 0 to 1 scale); and (4) GC content expressed as the percentage of G's and C's within the same 60-nt window as minimum free energy.

For the stoichiometry model, we fitted a linear model to the ΔCleavage-efficiency measurements of all the MAZTER-seq sites and m⁶A-seq sites using all variables described above, and performed a 5-fold cross-validation using in-house scripts to control for overfitting. For generating the graphical depictions of the code per nucleotide position the resulting model coefficients were scaled so that the average was 0.

For the binary (logistic) classifier, we modeled methylation fitting a logistic model assigning a positive value for all MAZTER-seq sites and m⁶A-seq sites, and a negative value for all the sites that were in confidence group 0 neither are m⁶A-seq sites. All variables aforementioned were used without performing feature selection and we performed a 10-fold cross-validation using in-house scripts to control for overfitting.

Evolutionary analysis

Pairwise alignments between the SK1 and the *S. mikatae* transcriptomes were performed as in (Schwartz et al., 2013). A custom script was used to identify all DRAC motifs occurring in either of the two transcriptomes, and extracting the sequence environment surrounding it. The binary classifier (above) was applied to each of the identified sites, and for each site the difference between the predicted scores in the two organism was calculated (Δ prediction). Sites that lacked a DRAC core (in one of the two organisms) were manually provided a predicted level of 0. Note that the logistic model was only trained on RRACA sites, yet it was applied to all DRAC sites, including both Ts at position -2 and nucleotides other than A at position +2. To expand the model, we treated Ts at position -2 as A (i.e., they received the same weight as an A at this position), and all positions at position +2 were treated as As (we found that

limiting ourselves only to RRACA sites did not alter our conclusions, and yet reduced the number of surveyed data points). For the calculation of the respective m⁶A levels in each of the two organisms, we utilized the available m⁶A-seq data that had been generated for both of them (Schwartz et al., 2013). For each site in each organism, we calculated the log2FCH in edgeR normalized coverage between IP and Input samples. The Δenrichment score was defined as the difference between this value in SK1 and the corresponding value in *S. mikatae*. For the analysis presented in Figures 4D and 4E, we limited ourselves to sites that had at most 2 mutations in the sequence window of 9 bp centered around the methylated site, and that had predicted levels of methylation greater than > 0.2 either in SK1 or in *S. mikatae*. In addition, we classified each site based on whether or not a ‘core’ consensus sequence - defined as presence of a DRAC motif - was present in each of the two organisms; This was used only for color-coding each point in Figure 4D.

Calculation of cleavage efficiencies at known sites in human and mouse samples

A catalog of m⁶A sites, divided into low, intermediate, and high confidence bins in mouse and human was obtained from Table S2 in (Schwartz et al., 2014a). For mouse, we intersected this dataset with a miCLIP dataset (Linder et al., 2015) obtained for mouse brain (A. Grozhik and S. Jaffrey, personal communication) and defined an ultra high confidence set of sites consistently identified in both (Table S5). In human, we intersected this database with a set of 81,520 sites collectively identified across six different miCLIP studies, and assembled as part of the WHISTLE database (Chen et al., 2019) (Table S6). These initial datasets were filtered to retain only (1) sites harboring an ACA motif, (2) sites within 40 bp or more from an upstream or downstream ACA sites (and hence quantifiable via MAZTER-seq), and (3) sites covered by at least 15 reads in at least one measurement. This setting for cleavage efficiency calculation was done using the mazter_mine pipeline (Figure S1), with the following non trivial parameters: bam2ReadEnds.R -l 300 -m 15; and master_mine.R -u 40 -d 40 for all datasets (Except METTL3-KO dataset, which used different parameters for the second step, being: -u 60 -d 60).

De novo detection of putative m⁶A sites in mouse

Putative m⁶A sites were identified using a comparison between normalized cleavage efficiency in WT mESC minus METTL3-KO mESC. Cross-sample quantile normalization was performed keeping only sites which were measured in at least one replicate of the two groups. Mean difference and p value of the t test were calculated for each site. For calling a site putatively methylated we defined a mean difference threshold to be the 1% quantile of the negative difference sites (~0.31 Cleavage Efficiency difference), p value thresholds were defined as p < 0.05 for confidence group 1, and p < 0.01 for confidence group 2 (group 1 estimated-FDR = 0.25, group 2 estimated-FDR = 0.11).

Statistical Analysis and plotting

All statistical analyses and visualizations were performed using the R language and the ‘base’ package (R Development Core Team, 2017), along with supplementary packages: ‘AUC’ (Ballings and Van den Poel, 2013), ‘dplyr’ (Wickham et al., 2015), ‘magrittr’ (Bache and Wickham, 2014), ‘tidy’ (Wickham and Henry, 2016), and ‘gtools’ (Warnes et al., 2014). Sequence logos were prepared using the ‘ggseqlogo’ package (Wagih, 2017), most plots were generated using the ‘ggplot2’ package (Wickham, 2010) along supplementary plotting packages: ‘RColorBrewer’ (Neuwirth, 2014), ‘LSD’ (Schwalb and Tresch, 2015), and ‘gridExtra’ (Augie, 2016).

DATA AND CODE AVAILABILITY

Raw data files have been deposited to the NCBI Gene Expression Omnibus under accession number GEO: GSE122961.

MAZTER-seq is a software we developed for analyzing MAZTER-seq derived data (Figure S1). It outputs cleavage efficiency calculations and quality control reports. The software along sample data is available through the GitHub repository https://github.com/SchwartzLab/mazter_mine.

Supplemental Figures

Cell

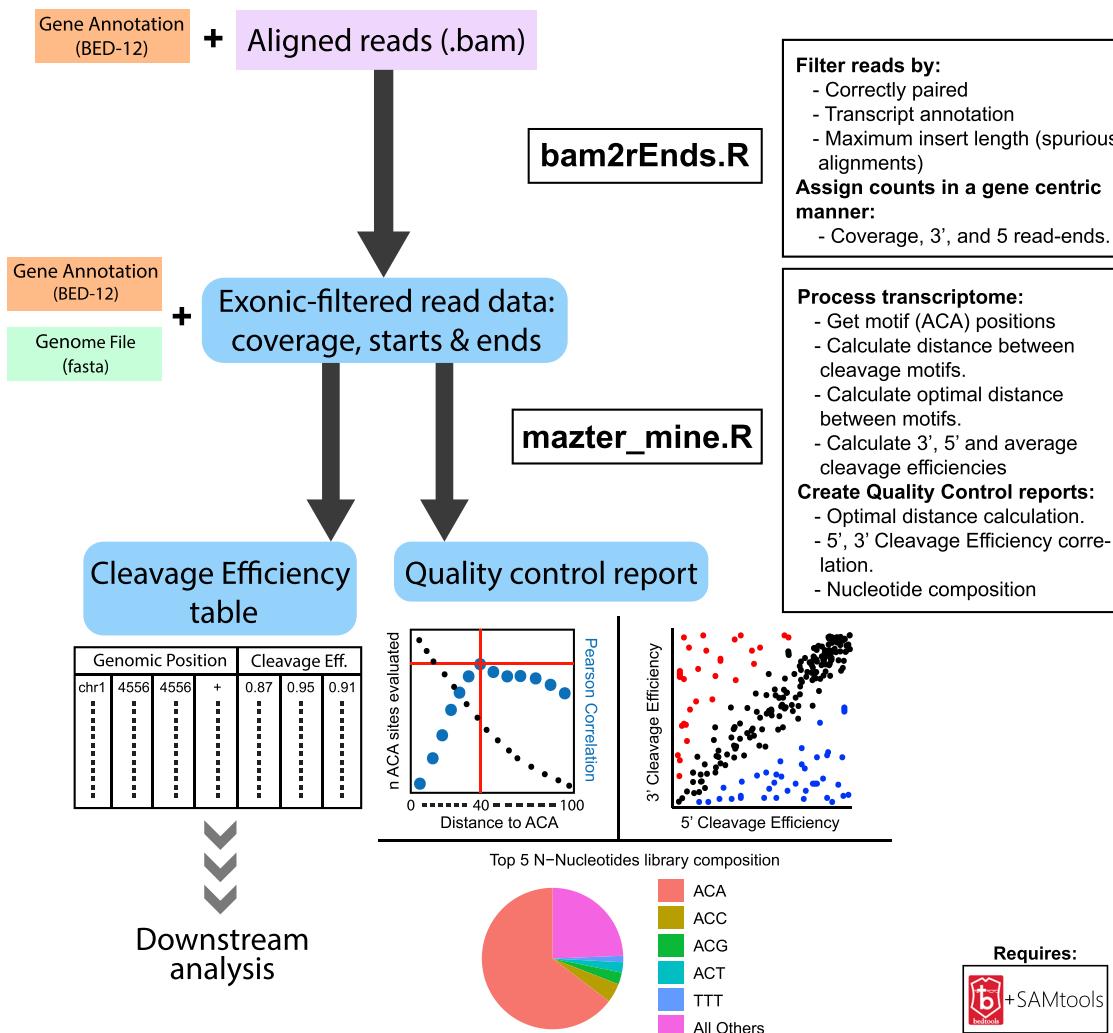


Figure S1. MAZTER-MINE, Related to STAR Methods

MAZTER-MINE receives as input an unsorted alignment file along with a genome sequence and transcriptome annotation. Per each genome, MAZTER-MINE will identify all ACA sites, and annotate them with the distances to the adjacent upstream ACA sites (in transcriptome space). Per each MAZTER-seq library, MAZTER-MINE will (1) identify all reads that are consistent with the transcriptome annotation, (2) identify the minimal distance to adjacent ACA sites that maximizes the consistency between the 3' and 5' estimates of cleavage efficiency (note that this is dependent on the insert size of the library), (3) generate quality control reports summarizing the fraction of reads beginning and terminating with ACA sites, and (4) generate transcriptome-wide cleavage efficiency tables for each quantifiable ACA site, along a QC report.

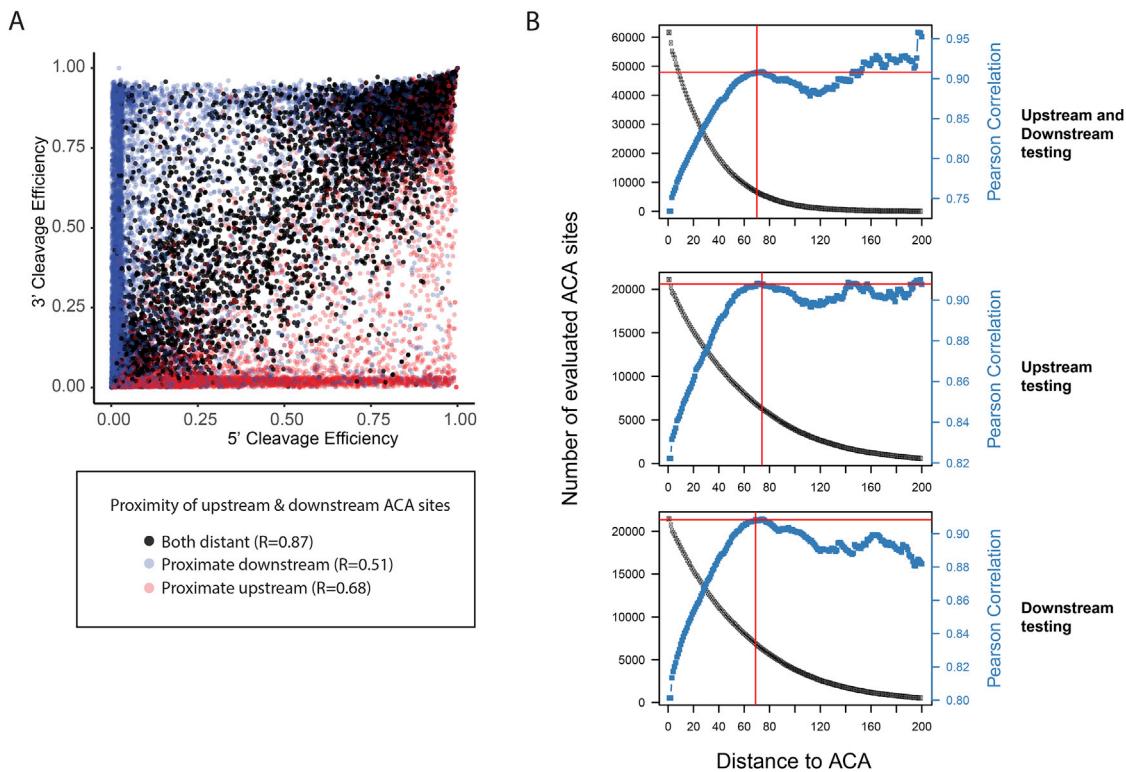


Figure S2. Quantitation of m^6A Signal Is Dependent on Distance to Adjacent ACA Sites, Related to Figure 1

Accurate quantification of 5' and 3' cleavage efficiencies is dependent, in a highly predictable manner, on the distance between the interrogated site and its nearest downstream and upstream ACA sites. (A) Scatterplot of 3' cleavage efficiency versus 5' cleavage efficiency. Points are color coded based on their proximity with other ACA sites: blue, if the plotted site is too close (< 60 nt) to a downstream ACA site; red if it is too close (< 60 nt) to an upstream ACA site; and black if it is ≥ 60 nt from both closest upstream and downstream ACA sites. (B) Minimal ACA distance calculation. Correlations between 3' and 5' cleavage efficiencies at different minimal ACA distances are shown in blue. The number of evaluated sites at each filtering step is plotted in black. Two red lines mark the point at which correlation stops increasing as minimal ACA distance increases.

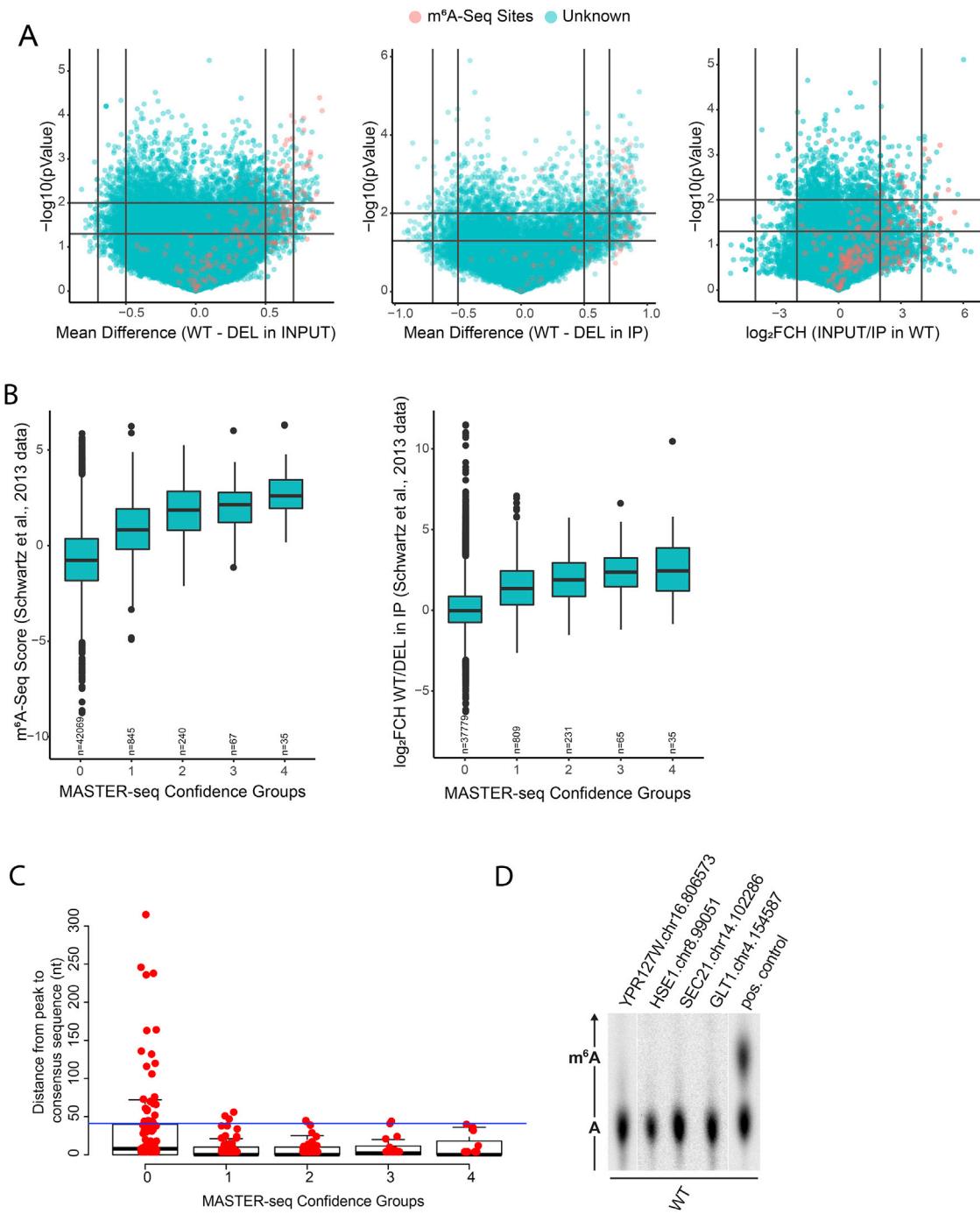


Figure S3. De Novo Detection of m⁶A Using MAZTER-Seq, Related to Figure 2

(A) Volcano plots of the three comparisons used in m⁶A sites *de novo* detection. Vertical and horizontal lines depict the scoring thresholds. x axis from left to right: 1) Mean difference between cleavage efficiency of each site in WT and IME4Δ/Δ strains in INPUT treatment, 2) Mean difference between WT and IME4Δ/Δ strains in m⁶A-IP treatment, 3) log₂-Fold-Change (Log2FCH) in Input over IP treatments in WT strain. Comparisons 1 and 2 thresholds are (left to right): -0.7, -0.5, 0.5 and 0.7; comparison 3 thresholds are: -4, -2, 2, and 4, respectively. y axis: -log₁₀ t tests p values; horizontal thresholds are (bottom-up): 0.05 and 0.01 for all plots. (B) Boxplots depicting the enrichment of m⁶A-seq derived scores (Schwartz et al., 2013) in MAZTER-seq derived confidence groups exclusively in MAZTER-seq *de novo* detected sites. (C) Boxplots depicting m⁶A-seq sites distance from peak to consensus sequence identified in (Schwartz et al., 2013) divided by MAZTER-seq confidence groups. A horizontal blue line marks the 41 nt distance threshold, and highlights that confidence group 0, in particular, is associated with a large number of sites that are not within vicinity of a near consensus site (24 out of 199 ~12%). (D) SCARLET analysis of four sites identified by m⁶A-seq, but with peaks > 100 bp from the nearest consensus sites.

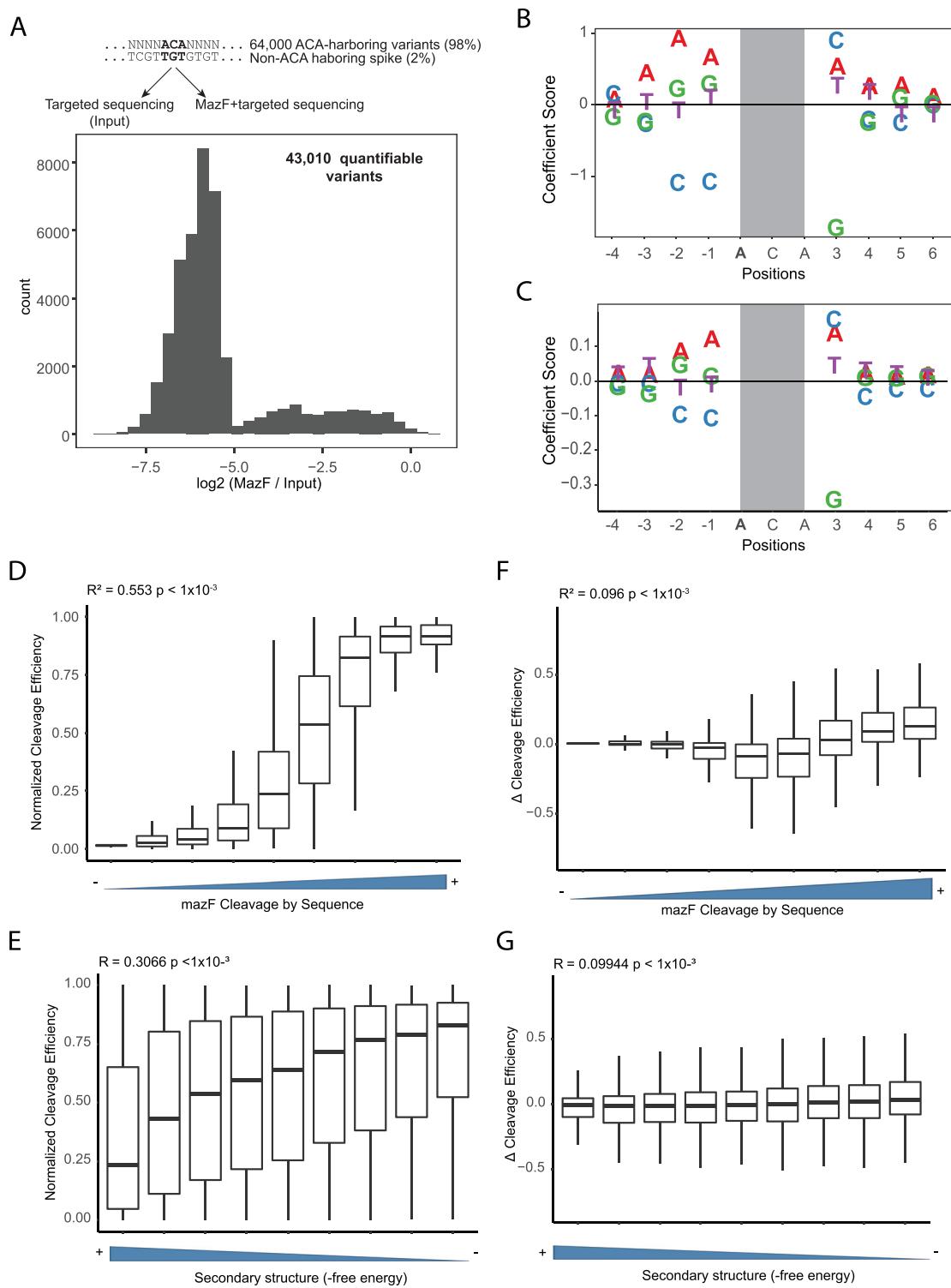


Figure S4. Establishment of Δ Cleavage Efficiency as a Readout, Related to Figure 3

(A) (Top) To directly assess the ability of MazF to cleave at different consensus sites, we in-vitro transcribed a synthetic pool of RNA containing thousands of sequence variants all harboring an NNNNACANNNN motif in the center. In parallel, a small amount (~2%) of a non-cleavable RNA spike (that did not have an ACA site) was added to this pool. The relative amount of each variant in this pool was quantified by targeted sequencing before and after cleavage by MazF. The relative abundances of each ACA containing probe were normalized by those of the non-cleavable spike, to yield an estimate of cleavage efficiency. (Bottom) Distribution of relative abundances following MazF cleavage. The vast majority of sites - regardless of the sequence preceding or following the ACA site - were

(legend continued on next page)

depleted in a highly efficient manner (A). Of particular importance, RRACA harboring sites were depleted, on median, 64 fold, and 98% of these sites were depleted at least 5-fold following cleavage, demonstrating that the overwhelming majority of sequences are cleaved in a highly efficient manner in the absence of methylation. (B) Based on the relative cleavage efficiencies derived from experiment in (A), a linear model was fitted, based on the identity of the nucleotides surrounding the methylation sites. The coefficients for each nucleotide at each position are plotted (as in Figure 3D). The y axis captures the extent to which a nucleotide contributes toward cleavage, with lower values being associated with decreased cleavage. The three features contributing toward cleavage efficiency are presence of C at position -1 and -2 reduced cleavage efficiency, as did presence of G at position +3. While presence of C at either position -1 or -2 is incompatible with a methylation consensus sequence and hence not expected to have a major impact our results, presence of G at position +3 is compatible with methylation. Importantly, these effects are quite mild: Even when G is present at position +3, RNAs harboring such Gs are depleted ~32-fold (median) from the original pool. Nonetheless, on the basis of these results we analyze and interpret sites harboring a G at position +3 with caution, in particular in analyses quantifying the extent to which m⁶A levels are predictable, or in analyses in which a methylation deficient control is lacking. (C) As a parallel approach to the one in panels A-B, cleavage efficiencies were modeled based on the genome-wide cleavage efficiencies across all ACA sites obtained in an IME4Δ/Δ strain ('cleavage model'). Coefficient plots for this model capture essentially the same features as in (B). (D) Depiction of the relationship between predictions made by the sequence based cleavage model (binned into equally sized bins) and between cleavage levels. (E) Correlation between secondary structure (estimated free energy) and cleavage efficiency. (F-G) As in D-E, but on the basis ΔCleavage efficiency.

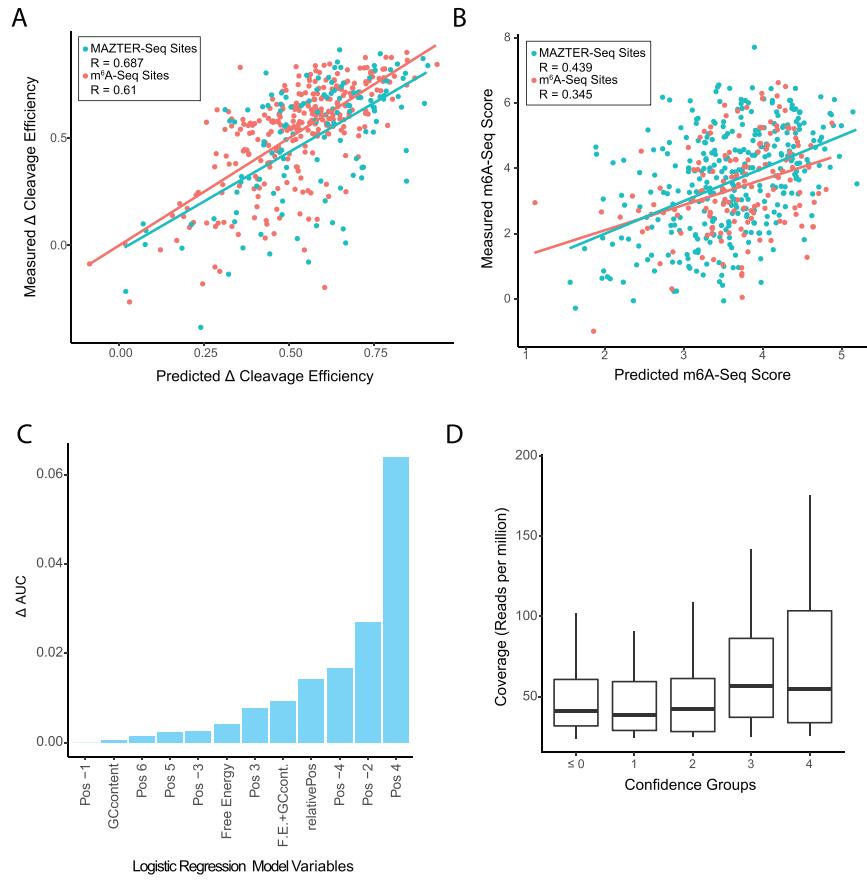


Figure S5. Characterization of m⁶A Code, Related to Figure 3

(A) Correlation between predicted and measured Δ Cleavage efficiency. Dots and regression lines are color-coded separately for MAZTER-seq and m⁶A-seq sites. (B) Correlation between predicted and measured m⁶A-seq score, color-coded as in (A). (C) Quantification of the relative contribution of each of the indicated variables to the performance of the logistic model. The bar plot depicts the difference in AUC from the full model when removing each of the variables in a 1-in-1-out fashion. (D) Coverage in reads per million by MAZTER-seq confidence groups.

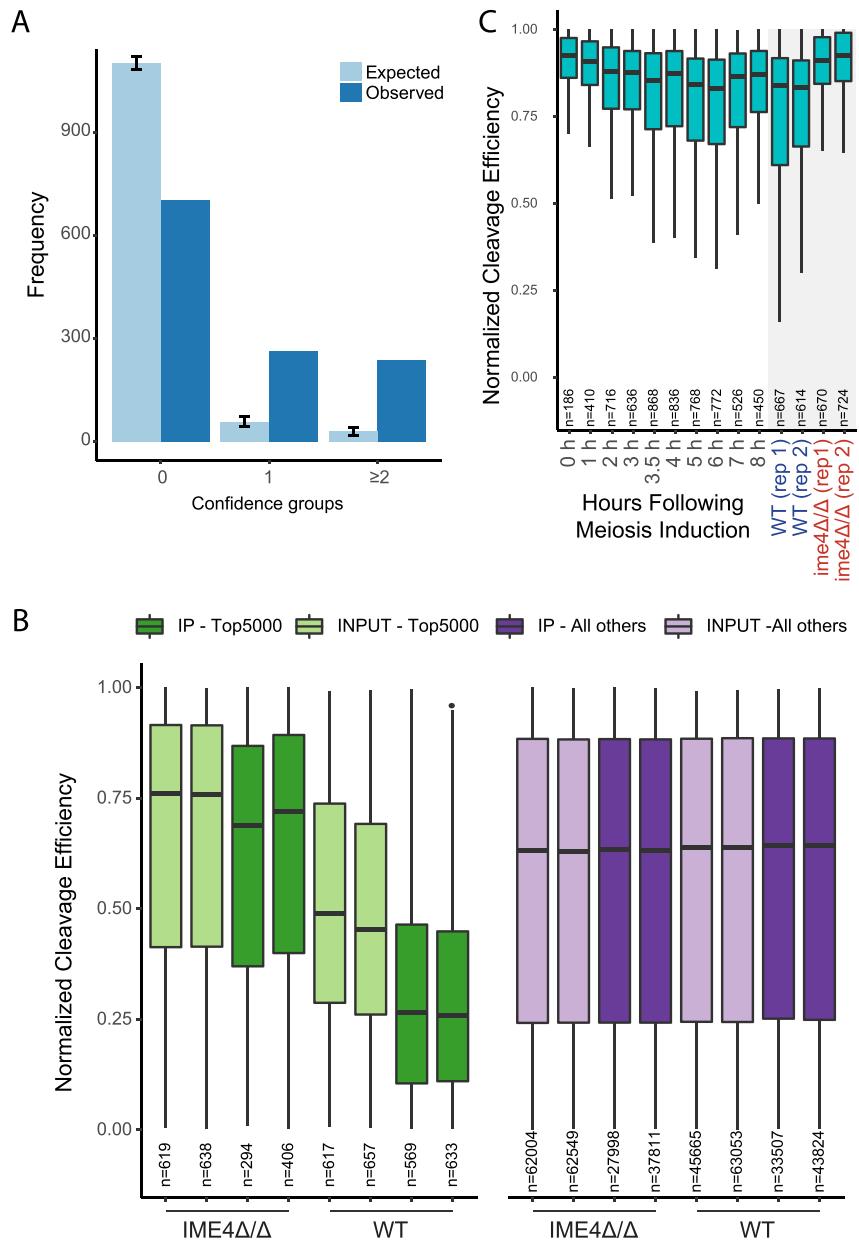


Figure S6. Top 5,000 Predicted Sites Characterization, Related to Figure 3

(A) Frequencies of expected and observed sites from different confidence groups in the Top-5000 sites set. Expected sites were calculated based on the frequency of each confidence in the entire set of sites (background). Interval bars represent 95% confidence intervals. (B) Normalized Cleavage efficiency distributions per sample. Top-5000 MAZTER-seq quantifiable sites are shown in green, and background levels of normalized cleavage efficiency are shown in purple. (C) Meiotic m⁶A dynamics in predicted sites measured with MAZTER-seq. Distributions of normalized cleavage efficiencies of Top-5000 predicted sites that do not overlap with any of the detected sites (either via MAZTER-seq or m⁶A-seq Sites).

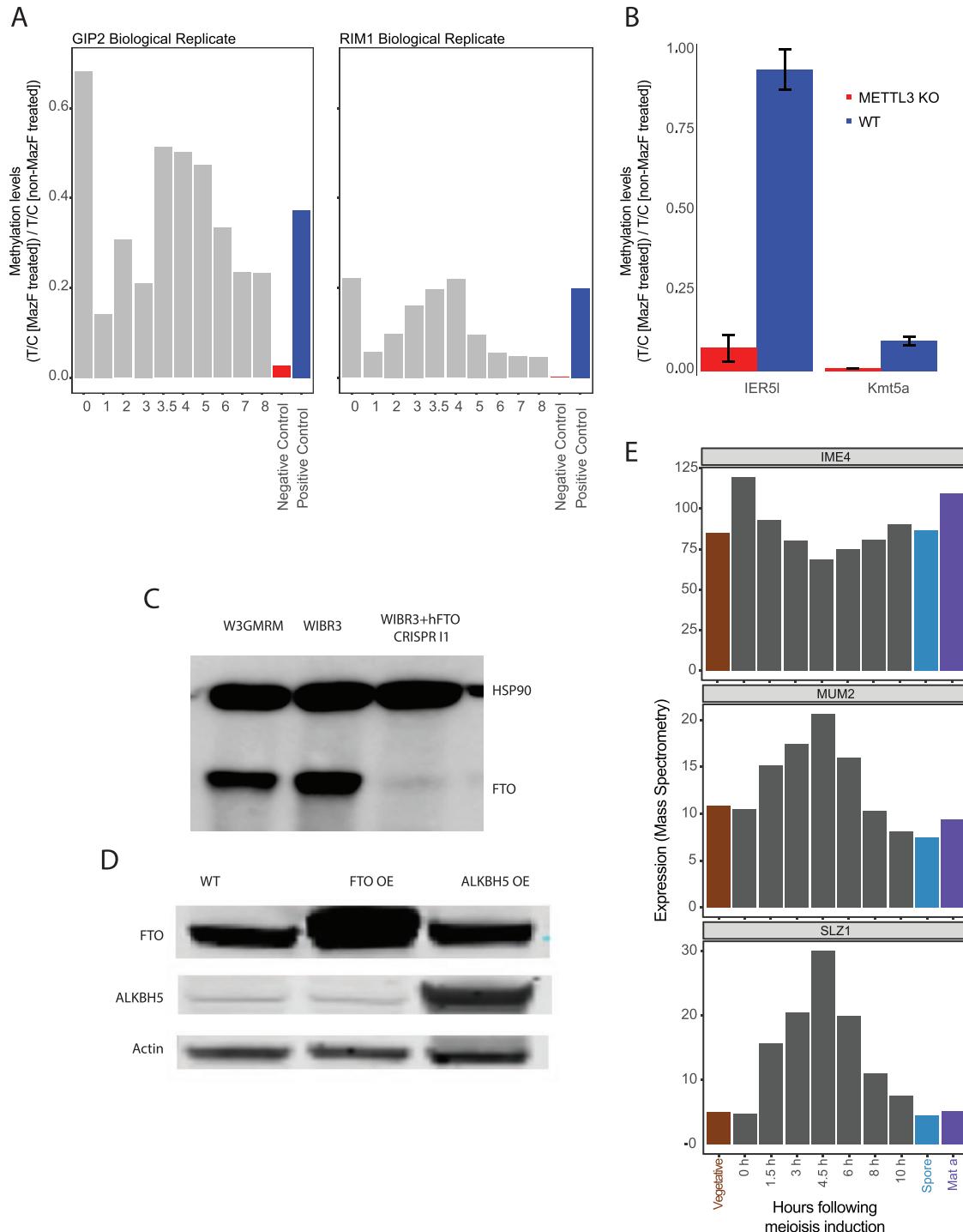


Figure S7. MazF qPCR Experiments, Strains Validation, and Mass Spectrometry of Meiosis Gene Markers, Related to Figure 5

(A) A biological replicate experiment, for MazF-qPCR on two targets across a meiosis time course (as in Figure 5E). Note that due to reasons unclear to us the 0 h time-point produced high levels of cleavage. This is likely an experimental artifact (as it is inconsistent with the MAZTER-seq results, and with the second replicate), but is included here for the sake of completeness. (B) MazF-qPCR on two target methylation sites in mouse, quantified either in mESCs or in METTL3 knockout counterparts. (C) Validation of FTO knockout in hESC cells, and (D) of FTO and ALKBH5 overexpression in HEK293T cells. (E) Mass spectrometry derived measurements of Ime4, Slz1 and Mum2 expression levels across a meiosis time course. Data obtained from (Cheng et al., 2018). Both Mum2 and Slz1 are highly induced at the onset of meiosis and peak at prophase likely underlying the temporal profiles of m⁶A along this time-course.